

OM of: US-09-303-518d-463 to: SPREMBL_19:* out_format: pfs
Date: Jun 30, 2002 8:07 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=framed_n2p.model -DEV=xlh
-O=cpn2.1/USPTO.SP00L/US093030518/runat.28062002.142714.4338/app.query.fasta.1.23501
-DB=SPREMBL_19 -QFWT=fastan -SUFFIX=rspt -GAPO=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPEXT=6.000 -GAPEXT=0.050 -YGAPEXT=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DEEXT=7.000 -START=1 -MATRIX=biom62
-TRANS=human40.cdi -LIST=100 -DOCALLIGN=200 -THR_SCORE=pct
-NORMEXT=HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US093030518 -ECN1_1.1967 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518d-463
Database: SPREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 883.700000

Sequence	Strd Orig	ZScore	Escore	len	Document
SP_bacteriap:09JVG1	+ 2200.00	2594.46	6.5e-137	556	Q9JVG1 neisseria meningitidis
SP_bacteriap:09K0F9	+ 1016.50	1196.95	6.1e-59	412	Q9K0F9 neisseria meningitidis
SP_bacteriap:09J3V9	+ 752.50	883.30	1.6e-41	467	Q9J3V9 neisseria meningitidis
SP_bacteriap:09K0F4	+ 740.50	868.45	1.0e-40	498	Q9K0F4 neisseria meningitidis
SP_bacteriap:09J6B3	+ 658.00	772.48	2.6e-35	422	Q9J6B3 neisseria meningitidis
SP_bacteriap:09JWK8	+ 450.00	524.47	1.4e-21	509	Q9JWK8 neisseria gonorrhoeae
SP_bacteriap:09JXD2	+ 414.50	483.01	3.1e-19	482	Q9JXD2 neisseria meningitidis
SP_plant:09PVB8	+ 390.00	453.99	1.3e-17	483	Q9PVB8 arabidopsis thaliana
SP_plant:09ZM08	+ 172.00	189.88	0.0035	894	Q9ZM08 arabidopsis thaliana
SP_invertebrate:09NMS7	+ 162.50	179.86	0.0127	894	Q9NMS7 caenorhabditis elegans
SP_invertebrate:020908	+ 162.00	179.54	0.0153	1079	Q20908 caenorhabditis elegans
SP_bacteriap:09A909	+ 161.50	182.69	0.0149	529	P9A909 microbacterium ammoniaphilum
SP_bacteriap:09CNE2	+ 158.50	159.27	0.0154	116	O9CNE2 pasteurella multocida
SP_bacteriap:09J552	+ 158.00	194.51	0.0432	3570	O9J552 homo sapiens (human)
SP_bacteriap:053478	+ 157.50	179.44	0.0262	456	Q53478 neisseria gonorrhoeae
SP_human:09U039	+ 157.00	171.47	0.0348	956	Q9U039 homo sapiens (human)
SP_human:09U040	+ 157.00	168.70	0.0376	1262	Q9U040 homo sapiens (human)
SP_human:09U048	+ 157.00	162.73	0.0444	2296	Q9U048 homo sapiens (human)
SP_human:09U035	+ 157.00	160.92	0.0467	2752	Q9U035 homo sapiens (human)
SP_human:09U082	+ 155.50	169.82	0.0436	944	O9U082 homo sapiens (human)
SP_human:076037	+ 155.50	168.80	0.0448	1045	O76037 homo sapiens (human)
SP_human:09GMD2	+ 155.50	168.23	0.0455	1107	Q9GMD2 homo sapiens (human)
SP_human:09GVM6	+ 155.50	167.79	0.0461	1157	Q9GVM6 homo sapiens (human)
SP_human:09G487	+ 155.50	167.43	0.0464	1187	Q9G487 homo sapiens (human)
SP_human:09H486	+ 155.50	167.43	0.0466	1199	Q9H486 homo sapiens (human)
SP_human:09H485	+ 155.50	167.31	0.0467	1214	Q9H485 homo sapiens (human)
SP_human:09H483	+ 155.50	167.30	0.0467	1215	Q9H483 homo sapiens (human)
SP_human:09H482	+ 155.50	166.97	0.0472	1256	Q9H482 homo sapiens (human)
SP_human:09H488	+ 155.50	163.23	0.0524	1827	Q9H488 homo sapiens (human)
SP_human:09H481	+ 155.50	161.76	0.0546	2167	Q9H481 homo sapiens (human)
SP_human:09N070	+ 155.50	161.52	0.0549	2169	Q9N070 homo sapiens (human)
SP_fungi:09J223	+ 154.00	172.20	0.0487	622	Q9J223 emmericella nidulans
SP_fungi:070495	+ 152.00	164.50	0.0744	961	O70495 mus musculus (mouse)
SP_human:09U036	+ 150.50	160.31	0.1091	1275	Q9U036 homo sapiens (human)
SP_human:015038	+ 150.00	156.97	0.1198	1783	O15038 homo sapiens (human)
SP_human:060382	+ 150.00	156.92	0.1200	1791	O60382 homo sapiens (human)

SP_human:043418	+ 149.00	168.21	0.0985	513	Q43418 homo sapiens (human)
SP_human:09N0E4	+ 148.00	163.70	0.1258	716	Q9N0E4 homo sapiens (human)
SP_bacteriap:09Z4N7	+ 148.00	154.11	0.1646	1873	Q9Z4N7 enterococcus faecalis
SP_mammal:029071	+ 145.50	163.78	0.1689	528	Q29071 sus scrofa (pig)
SP_invertebrate:09U043	+ 145.00	157.29	0.2250	954	Q9U043 drosophila melanogaster
SP_invertebrate:09P0P2	+ 145.00	149.78	0.2652	2025	Q9P0P2 mus musculus (mouse)
SP_bacteriap:085018	+ 144.00	154.48	0.2618	1122	O85018 mycobacterium smegmatis
SP_invertebrate:063661	+ 143.50	149.15	0.3225	1805	Q63661 rat mus musculus (mouse)
SP_human:09N0M8	+ 142.00	150.66	0.3694	1299	Q9N0M8 homo sapiens (human)
SP_invertebrate:076894	+ 142.00	147.43	0.4043	1795	Q76894 drosophila melanogaster
SP_invertebrate:09VXMS	+ 142.00	144.51	0.4387	2406	Q9VXMS drosophila melanogaster
SP_invertebrate:09J0670	+ 141.50	144.87	0.4609	2187	P70670 mus musculus (mouse)
SP_human:014879	+ 141.00	158.17	0.3371	533	Q14879 homo sapiens (human)
SP_human:060585	+ 141.00	154.06	0.3782	830	O60585 homo sapiens (human)
SP_human:09H306	+ 140.50	152.80	0.4157	877	Q9H306 homo sapiens (human)
SP_human:09H307	+ 140.50	152.79	0.4158	878	Q9H307 homo sapiens (human)
SP_human:014651	+ 140.50	151.93	0.4259	957	O14651 homo sapiens (human)
SP_human:09JWK9	+ 140.50	149.53	0.4555	1217	Q9JWK9 homo sapiens (human)
SP_human:09G012	+ 139.50	150.40	0.5006	1236	Q9G012 homo sapiens (human)
SP_bacteriap:09CNC9	+ 138.00	147.64	0.5334	109	O9CNC9 pasteurella multocida
SP_human:09N0R59	+ 138.00	145.82	0.6788	1312	Q9N0R59 mus musculus (human)
SP_invertebrate:09Z0R4	+ 138.00	143.16	0.7324	1714	O9Z0R4 mus musculus (human)
SP_invertebrate:0952Z7	+ 137.50	146.85	0.7010	1116	Q952Z7 caenorhabditis elegans
SP_bacteriap:09JLH1	+ 137.00	159.39	0.5237	299	Q9JLH1 bifidobacterium animalis
SP_invertebrate:019930	+ 137.00	149.01	0.7002	847	Q19930 caenorhabditis elegans
SP_invertebrate:09W406	+ 136.50	161.07	0.5301	238	Q9W406 drosophila melanogaster
SP_invertebrate:0960E9	+ 136.50	152.29	0.6777	574	O960E9 drosophila melanogaster
SP_invertebrate:09V515	+ 136.50	149.68	0.7291	746	O9V515 drosophila melanogaster
SP_virus:09Q0P87	+ 136.50	147.96	0.7650	886	Q9Q0P87 human herpesvirus 4
SP_human:09Q2954	+ 136.50	143.37	0.8698	1404	Q9Q2954 human herpesvirus 4
SP_virus:056854	+ 136.00	147.72	0.8174	856	O56854 human herpesvirus 4
SP_fungi:09P3J0	+ 136.00	136.67	1.11	2592	Q9P3J0 neurospora crassa
SP_plant:020354	+ 135.00	154.07	0.7705	402	O20354 arabidopsis thaliana
SP_fungi:007229	+ 135.00	147.00	0.9391	817	O07229 saccharomyces cerevisiae
SP_invertebrate:044548	+ 135.00	146.29	0.9578	877	Q44548 caenorhabditis elegans
SP_human:09H7M4	+ 135.00	142.04	1.08	1343	Q9H7M4 homo sapiens (human)
SP_invertebrate:088491	+ 135.00	133.50	1.30	2588	O88491 mus musculus (mouse)
SP_invertebrate:09W2E1	+ 134.50	140.49	1.20	1479	Q9W2E1 drosophila melanogaster
SP_invertebrate:022579	+ 134.50	138.29	1.27	1844	Q22579 caenorhabditis elegans
SP_invertebrate:017921	+ 134.00	146.42	1.07	769	Q17921 caenorhabditis elegans
SP_bacteriap:088093	+ 134.00	140.61	1.26	1377	O88093 escherichia coli
SP_bacteriap:047692	+ 134.00	139.04	1.32	1611	O47692 escherichia coli
SP_fungi:09P038	+ 133.50	143.07	1.25	1014	O9P038 cyrtosia sativa (rice)
SP_invertebrate:057880	+ 133.50	151.17	1.30	1151	O57880 gallus gallus (chicken)
SP_fungi:09B7107	+ 133.00	144.25	1.29	849	Q9B7107 saccharomyces cerevisiae
SP_bacteriap:09B621	+ 133.00	140.88	1.41	1190	Q9B621 homo sapiens (human)
SP_invertebrate:09BKV7	+ 133.00	139.81	1.46	1325	Q9BKV7 leishmania major
SP_human:09V474	+ 133.00	137.59	1.55	1656	Q9V474 homo sapiens (human)
SP_invertebrate:09W204	+ 132.50	140.38	1.52	1179	Q9W204 drosophila melanogaster
SP_human:09Q945	+ 132.00	139.77	1.66	1182	Q9Q945 homo sapiens (human)
SP_virus:09DWM8	+ 132.00	139.29	1.64	1240	Q9DWM8 rat cytomegalovirus
SP_invertebrate:09W0B0	+ 132.00	137.87	1.73	1430	Q9W0B0 drosophila melanogaster
SP_invertebrate:09H056	+ 131.50	144.48	1.53	694	Q9H056 halobacterium salinarum
SP_bacteriap:09B349	+ 131.50	137.46	1.86	1404	Q9B349 homo sapiens (human)
SP_human:09P166	+ 131.50	137.92	1.94	1654	Q9P166 homo sapiens (human)
SP_mammal:018984	+ 131.00	147.59	1.47	460	O18984 cercopithecus aethiops

seq_name: SP_bacteriap:09JVG1

seq_documentation_block:

AC 09JVG1 PRELIMINARY;

DT 01-OCT-2000 (TREMURel. 15, Created)

DT 01-DEC-2000 (TREMURel. 15, Last sequence update)

DE PUTATIVE ADHESIN.

GN MAFB3 OR NMA0853.

OS Neisseria meningitidis (serogroup A).

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=56599;

480 rgtYrIleProProLysAsnTYrSerProSerAlaProLeuProLysGly 496
1443 1443
497 ProAsnAsnGlyTYrLeuAspLysPheGlyAsnGlyTrpThrLysGlyPr 513
1444 .AGAAGACTCAGAGTACAGTAAAGCCATCGCA..... 1482
513 oSerArgThrLysGlyGlnGlnPheGlyTrpAspValGlnLeuSerLysT 530
1483CGAGA.....TGGCAATAAACAAGGTTAGATTAT 1518
530 hGlyArgGlnGlnLeuGlyTrpAlaSerArgAspGlyLysHisLeuAsn 546
1519 CATTATAGTGTCGATATCAATAGAA 1548
547 lIleSerIleAspGlyLysIleThrHisLys 556

seq_name: sp_bacteriap:Q9KOF9

seq_documentation_block:

ID Q9KOF9 PRELIMINARY; PRT; 412 AA.
AC Q9KOF9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE HYPOTHETICAL PROTEIN NMB0647.
GN NMB0647.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID:491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MC58 / SEROGROUP B;
RX MEDLINE-20175755; PubMed-10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
Rosen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
Nelson W.C., Galin M.L., Deboy R., Peterson J.D., Hickey E.K.,
Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
Cotton M.D., Ueberback T.R., Khouri H., Olin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002419; AAF41068.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 412 AA; 45508 MW; 6B41BAD5323C9052 CRC64;

alignment_scores:

Quality: 1016.50 Length: 449
Ratio: 3.592 Gaps: 12
Percent Similarity: 63.029 Percent Identity: 51.448

alignment_block:

US-09-303-518D-463 x Q9KOF9

Align seg 1/1 to: Q9KOF9 from: 1 to: 412

588 CACCAAGAGATGAGCGAGGATTT.....CAAAAGCG 619
||| |||
42 AspLysGlnThrAlaArgLysIleTYrSerValIleLysGlnGlnValVal 58
620 CCACCCGATACAGCCCGAGCTGGACAGATCGGGCAATGCCGCCGAGCC 669
||| |||
58 shisGlnCysSer...SerThrGlyArgIleThrGlnCysArgGlnAsnI 74
670 TTCAAGCGGACATGACATATCTCAAAACATCATCGCGCGGACAGAGAG 719
::: ||| |||
74 legIatArgIleIleGlnPheThrGlnAsp..... 83

720 AATTGTGGCGGACGCGATCGCTGCAGGGTATTAAGCAAGCTCAACA 769
::: ||| |||
84LysArgPheAspSerArgPheLysAspLeuLysGly 96
770 TTGCTGATGACACGCGCTTGCTGCTTCCACCAACCAAGATGGCG 819
: |||
96 u.....SerLeuTYrTYrLeuAsnLysHisPro 105
820 CGCATCAACGATTTGCAGATATGCGCACTCAAGACTATGCCGACG 869
||| |||
106AspLeuValAlaSerTYrLeuLysAlaGlnTYrIleLysLe 119
870 AGCCATCCGCGATTTGGCGACGCCCAAAACCCCAATGGCCGACAGCA 919
::: ||| |||
119 uAspArgGlnAspLysSerIleLeuHisArgTYrIleSerProGlyAla 136
920 AAGCCGTCAGC...AATATCTTATGGCAGCCATCCCATCAAGGAT 966
|| |||
136 lIleValIleSerGlySerLeuGlyValValLeuSerGlyValAlaGly 152
967 GAGCTGTCCGGGAAATACGGCTTGGC.....GG 998
||| |||
153 GlySerCysAlaGlnThrPheGlyLeuGlyCysAlaAlaLeuValGln 169
999 CATCAGCGCA.....CATCTGTCAAGCGGTGCGAGATGGCGCGA 1039
||| |||
169 yValThrSerSerTYrAspHisValIleThrGlyThrLys..... 182
1040 TGCATTCGCGAAGGAATCCGCGCTCAGCGACATTTGCCGATGCG 1089
183 |||
1090 GCATACGCCAATACCCGCTCCCTTACATCCCGAATATCGCTCAAA 1139
||| |||
188 AlaSerGlnGlnArgPro..... 193
1140 CTGGAGGACGCGTTAGCGCAAGAAACATCACTCTCTCAACCGTCCG 1189
194 ThrIleAlaVal 197
1190 CGTCAACGCGCAAAATGTCAAACTGGCAGCAACGCCGCGAAGACA 1239
::: ||| |||
197 alGlnAlaLeuLysGlnLeuGlyLeuSerGlnGlnAla..... 209
1240 GCGGTACGCTTTGACGCTAAGAGGTTCCGAATTTGAGAACGCTGAA 1289
210 AlaGlnTYrValGln 214
1290 ATATGATACGAAGCTCGATATTCAGAAATTATCGGGGCGGTATACCTA 1339
::: ||| |||
214 nPheSerIleAspLeuPheSerValGlnLysSerGlyGlyIleProL 231
1340 AGGCTAAGCCTGTGTTGATGCGAAACCGAGATGGAGGTGATAGAG 1389
||| |||
231 ySalAlaLysProValPheAspAlaLysProArgTrpGlnValAspArgLys 247
1390 CTTAATTAATTTACAACCTCGTGCAGAGGTGGAGAAAATTTTCAGAAAC 1439
||| |||
248 LeuAsnLysLeuThrThrArgGlnGlnValGlnLysAsnValGlnGln 264
1440 GAGAGAGAGAGTACAGTACAGTATTAAGCCCATCGGACAGAGAT 1489
||| |||
264 rArgArgArgSerGlnSerSerGlnPheLysAlaHisAlaGlnArgGln 281
1490 GGGAAATTAACACAGGTTAGATTTAATCAATTTATAGTGTGATATC 1539
||| |||
281 rGlnAsnLysThrGlyLeuAspPheAsnHisPheIleGlyValAspLe 297
1540 AATAGAAAGCAGACAGTAAACAGAGGCGATAGTCAACCGGTGATGT 1589
||| |||
298 AsnLysLysGlyThrValThrGlyGlnHisSerLeuThrArgLysAsp 314

320 320
1001 TCACGGCAGATCTCTCAAGGGGTCGACGATGGCGGATGCTTCCG 1050
321 LysValAlaLysLeuAlaLysAlaLysPro 331
1051 AAGGGAATCCGCCGTCAGCAGCAATTTCCGATGCGGATACGCCAA 1100
332 ...GlyLysAlaLysAlaLysSerGlyAspPheAlaLysSerTyrLysLys 347
1101 ATACCCGCCCCCTTACCATTCGCCAAATATCCGTTCAAACTTGAGCAGC 1150
347 sLeuAlaLeuSerAspSerAlaArgGlnLeuTyrGlnAlaLysTyr 364
1151 GTTACGGCAAGAAACATCATCTCTCAACCCGCGCGCTCAAGGC 1200
364 rg.....GluAla 366
1201 AAAAATGCAAACTGCGACAGCAACGCCCGAAGACAGCGCTACCGTT 1250
367 LeuAspLeuHisTyrGlnAspLeuLeuArgArgLysThr..... 379
1251 TGACGGTAAGGGCTTCCGAATTTGGAAGCAGCAGCTAAATATGATGCA 1300
380 .AspGlySer.....LysPheLLeuAsn 388
1301 AGCTCATATTCAAGAATTATCGGGGGCGGTATACCTAGGCTAACGCT 1350
388 LysArgGlnLeuAspAlaValThrAsnAspAlaLeuLeuGlnAlaLysArg 404
1351 GTGTTTATGCGAAACCCGAGATGGAGCTTGATAGAACTTAATAATT 1400
405 ThrLeuSerAla.....LeuAspLysProLysAsnPhe 416
1401 GACA.....ACTGTGAGCAGGAGGAGAAATTTTACAGAAACAGAA 1444
416 uAsnGlnLysAsnArgLysGlnLeuLysAlaThrLeuGlnAlaLysAsn 433
1445 GAAAGACTCAGAGTACTGCTT.....AAGCCCATGCGCAA 1482
433 LngLngLysArgAlaGlnPheThrPheLysTyrGlnLysHisSerGln 449
1483 CGAGAA.....TGGCAAAATTAACAGGCTTA 1509
450 ValLysSerTyrLeuLysSerLysGlyLysLe 460
seq_name: sp_bacteriap:09JSV9
seq_documentation_block:
ID 09JSV9 PRELIMINARY; PRT; 498 AA.
AC 09JSV9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ADHSIN MAFB2
GN MAFB2 OR NMA2113
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parthill J., Achman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Daves R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,
RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rastandean M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.,
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506(2000).
DR EMBL; AL162758; CAB85327.1; -

KW Complete proteome.
SQ SEQUENCE 498 AA; 54653 MW; 38F1E7C12B8811F7 CRC64;
alignment_scores:
Quality: 740.50 Length: 586
Ratio: 2.165 Gaps: 18
Percent Similarity: 58.362 Percent Identity: 32.253
alignment_block:
US-09-303-518D-463 x 09JSV9 ..
Align seg 1/1 to: 09JSV9 from: 1 to: 498
28 ATTGTGTCATACCTGCGACGTGTC.....CTGCCGAT 59
7 LeuLeuLeuLeuAlaLysAlaLysAlaLysAlaLysAlaLysAlaLysLeu 23
60 GCATGACACAGCCCTCAGATTGGCAACAGATCCCTTTATCCGCGAGCTTC 109
23 nProAlaLeuAlaLysAlaLysAlaLysAlaLysAlaLysAlaLysAlaLys 40
110 TCGACCGTCAGCATTTGCAACCCGAGCGGAATACCATTCATTCGCGAC 159
40 LngLysArgGlnHisTyrGlnProGlyLysTyrHisLeuPheLys 56
160 ...AGGGGGAGCTTGCAGGCGCAAGCGCATATCGGATTGGGAACAT 206
57 ProArgGlySerValSerAspArgThrGlyGlnLeuAlaLysAlaLys 73
207 ACAAGCCATCAGTTGGGCGCCTGATGATGATCAACAGCGCGCGCTGAG 256
73 pTyrThrHisArgMetGlyAsnLeuLeuLeuGlnAlaLysAlaLys 90
257 GAAATATCGCTACATTTGCGCTTTTCCGATCAGCAGGCGCAAAATTCAT 306
90 LysThrLeuLeuLysHisThrArgPheSerGlyHisGlyTyrGlnLys 106
307 TCGCCCTTCGACACCATATGCTCAGATTCGATTTGTCAGAAACCGCTAG 356
107 AlaProPheAspAsnHisAlaLysAspSerAlaSerGlnGlnLysLys 123
357 TCCCGTTACGAGTATCAGCTTTTACCGATGATGATGATGATGATGATGAT 406
123 nValAspLeuLysGlnPheThrValTyrArgLeuAsnTrpLeuGlnHisGln 140
407 ACCATCCCGCGAGCGGTATGACAGGCGCCACAGGGCGCGCTATCCGCT 456
140 LysHisProAlaAspAlaLysArgLysProLysGlyLysAsnTyrProLys 156
457 CCCAAAGCGCGAGGATATATACAGCTAGCATTAAGCGCTTGCCCA 506
157 ProThrGlnAlaLysArgAspGlyTyrThrTyrHisValAsnGlyThrAla 173
507 AAATATCCCGCTCAACCTGAGCAACCGCAACCGGACCGGACCGGCTG 556
173 gSerLeuLysLeuAsnProThrAspThrArgSerTyrLeuArgGlnArgLys 190
557 CCGACCGTTTCCACATGCCCGCGCTATGCTGACGCAAGAGTARGGCGAC 606
190 eAspAsnTyrSerAsnLeuLysSerAsnPheSerAspArgAlaLysPro 206
607 GGATTCAACAGCGCGCCAGCATACAGCCCGAGCTGACAGATGGGCGAA 656
207 AlaAsnArgLysMetPheGlnHisAsnAlaLysLeuAspArgTyrGlyAs 223
657 TGGCGCGAGCGCTTCAACGAGCACTGCAAGATATGCTCAAAATCATCTCG 706
223 nSerMetGlnPheHisGlnGlyValAlaLysAlaLysAlaLysAlaLys 240
707 GCGCGCGAGGAAATTTGCGCGCGAGCGATGCGCGAGGATTAAGC 756
240 LeuSerAlaLysGlnAlaLysGlnLysLeuLysAlaLysLeuTyrGlnThrArg 256

757	GAAGCGCTCAAACTGGTCGTCATCCACAGCGCTGGCTGCTGCTTCACACGA	806
758	807
759		808
760		809
761		810
762		811
763		812
764		813
765		814
766		815
767		816
768		817
769		818
770		819
771		820
772		821
773		822
774		823
775		824
776		825
777		826
778		827
779		828
780		829
781		830
782		831
783		832
784		833
785		834
786		835
787		836
788		837
789		838
790		839
791		840
792		841
793		842
794		843
795		844
796		845
797		846
798		847
799		848
800		849
801		850
802		851
803		852
804		853
805		854
806		855
807		856
808		857
809		858
810		859
811		860
812		861
813		862
814		863
815		864
816		865
817		866
818		867
819		868
820		869
821		870
822		871
823		872
824		873
825		874
826		875
827		876
828		877
829		878
830		879
831		880
832		881
833		882
834		883
835		884
836		885
837		886
838		887
839		888
840		889
841		890
842		891
843		892
844		893
845		894
846		895
847		896
848		897
849		898
850		899
851		900
852		901
853		902
854		903
855		904
856		905
857		906
858		907
859		908
860		909
861		910
862		911
863		912
864		913
865		914
866		915
867		916
868		917
869		918
870		919
871		920
872		921
873		922
874		923
875		924
876		925
877		926
878		927
879		928
880		929
881		930
882		931
883		932
884		933
885		934
886		935
887		936
888		937
889		938
890		939
891		940
892		941
893		942
894		943
895		944
896		945
897		946
898		947
899		948
900		949
901		950
902		951
903		952
904		953
905		954
906		955
907		

```

1621 AAACATGGGGCTTTATCAAGCGAAGTGGAAATTAATAAAGCCTGATGGAAG 1670
      :::: ||::: |||||
465 ATGSPGilyIleArgTleArgThrVal..... 473
      :::: ||::: |||||
1671 TTGGGAGGTGGAAACGAAAAAAAGTGGGAAAGTGATGACCAAGCACACA 1720
      ::::| | | | | |||||::: ||
474 .TyrGIuPProAlaThr.....GlyIysValVal.....Thra 484
      ::::| | | | | |||||::: ||
1721 TGTTCACA 1728
      |||||
484 lAphePro 486
      |||||
seq_name: sp_bacteriap:Q9K0F4

```

ID	seq_documentation_block:	AC	Q9K0F4	PRELIMINARY:	PRT:	422 AA.
AC	Q9K0F4;					
DT	01-OCT-2000 (Tremblrel, 15, Created)					
DT	01-OCT-2000 (Tremblrel, 15, Last sequence update)					
DT	01-DEC-2001 (Tremblrel, 19, Last annotation update)					
DE	MAPB-RELATED PROTEIN.					
CN	NMB0653.					
OS	Neisseria meningitidis (serogroup B).					
OC	Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.					
OX	NCBI_TaxId=491;					
RN	[1]					
RP	SEQUENCE FROM N.A.A.					
RC	STRAIN-MC58 / SEROGROUP B;					
RX	MEDLINE=20117575; Pubmed=10710307;					
RA	Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,					
RA	Eisen J.A., Kechum K.A., Hood D.W., Peden J.F., Dodson R.J.,					
RA	Nelson W.C., Gwyn M.L., Debroy R., Peterson J.D., Hickey E.K.,					
RA	Halt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,					
RA	Mason T.J., Clecko A., Parsey D.S., Blair E., Clittone H., Clark E.B.,					
RA	Cotton M.D., Ulfersdal V., Khouri H., Qin H., Vamathevan J.,					
RA	Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,					
RA	Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;					
RT	Complete genome sequence of Neisseria meningitidis serogroup B strain					
RT	MC58 ⁺ .					
RL	Science 287:1809-1815(2000).					
DR	EMBL; AE002420; AAF62314.1; -.					
DR	TIGR; NMB0653; -.					
KW	Complete proteome.					
SQ	SEQUENCE 422 AA; 45845 MW; B5428F16A296585F CRC64;					

[illegible]

alignment_block:

Align seg 1/1 to: Q9K0F4 from: 1 to: 422

[illegible]

```

71 eGlnAspTyrThrHisGlnMetGlyAsnLeuLeuIleGlnGlnAlaAsnI 88
251 TTGAAGAAATATCGGTACATCTTGCCGCTTTCCGATTCACGGCCACAAA 300
88 IeAsnGlyThrIleGlyTyrHisThrArgPheSerGlyHisGlyIle 104
301 TTCCATTCGCGCTTGACACACATCGCTACATTCGATTCGATTCGACGAG 350
105 GlnHisAlaProPheAspAsnHisAlaAlaAspSerAlaSerGlnGly 121
351 CGGATGCGCTTGACGATTCACGCTTTACCGCATCCATTCGATTCGACGAT 400
121 sGlyAsnValAspGlnGlyPheThrValTyrArgLeuAsnTrpIleGly 138
401 AGCAACACATTCGCGCGACGCTATGACGCGGACGACGCGGCGGCTAT 450
138 IsglnHisHisProAlaAspAlaIleTyrAspGlyProGlyGlyAsnTyr 154
451 CCCGCTCCCAAGGCGGAGGATATACAGCTACGACATAAAGGCGT 500
155 ProGlyProThrGlyAlaArgAspGlyTyrThrTyrHisValAsnGly 171
501 TGCCCAAAATATCGCGCTCAACCTGACCGACGACGACGACGACGAC 550
171 TAlaArgSerIleLeuAsnProThrAspThrArgSerIleArgGlnAla 188
551 GGCTTGGCGACGCTTTCACATGCGCGCTATGCTGACGACGACGAGTA 600
188 rGlyLeuSerAspAsnTyrSerAsnLeuGlySerAsnPheSerAspAla 204
601 GGCACGCGATTCAAACGCGCGACCGCATACAGCCGCGAGTGCAGCATC 650
205 AspGlnIleAsnArgGlyMetPheGlnHisAlaIleGlyLeuAspArg 221
651 GGGCAATGCGCGGAGCGCTTCAACGCGCATCGCATATGCTCAAAAACA 700
221 pGlyAsnSerMetGlnPheIleAsnGlyAlaIleAlaIleAlaIleAla 232
701 TCATCGCGCGCGACGAGAAATGTCGCGCGCGCGCGCGCGCGCGG 750
233 .AlaGlyAlaLeuAsnProPheIleSerIleGlyAlaValAlaAsp 247
751 ATAAAGCAAGGCTCAACATTCGTCATGACGCGCTTGCGTCTGCTTTC 800
247 ..... 247
801 CACCGAAACAGATGGCGCGCATCAACGATTTGGCAGATATGCGCCAC 850
247 ..... 247
851 TCAAAAGCTATGCGCGACGACCATCCGATTCGCGAGTTCGCAAAACCC 900
248 ..... GlnTrpMetGlnGlnAsnPro 254
901 AATCGCGCAAGCATAGACGCGTCACGATATCTTTATGGCAGCCAT 950
255 AsnAlaIleGlnThrValGlnAlaLeuValAsnValLeuProPheAla 270
951 CCCCATCAAGGATTCGAGCTGTCGCGGAAATACGCGCTTGCGGCGGA 1000
270 ..... 270
1001 TCACGCGCATCTCTCAAGCGGTGCGAGTGGCGCGCATTCGATTCGCG 1050
271 ..... LysValIysAsnLeuThrLysAlaIleAlaLysPro 281
1051 AAAGGAAATCCGCGGTGCGAGCATTTTGGCGATGGCGCATACGCCAA 1100
282 ...GlyLysAlaIleValSerGlyAspPheSerAspSer..... 293
1101 ATACCGGTCCCTTACCATTCGCCGAAATATCCGTTCAAACTTGAGACGC 1150
.....

```

```

294 ..... TyrLysHisAsnThrAlaSerArgLeuSerGln 305
1151 GTTACGGCAAGAAACATCACCTCTCAACCGTCCGCGCGTCAACGCGC 1200
305 eValAspGlyGlnMetPheGlnThrArgAsnVal.....AspPhe 318
1201 AAAAATGTCAAACTGGCAGACCCACGACCCGCAAGACAGCGGTACCGT 1250
319 LysAlaLysSerIleGlyThrLysIleHis..... 328
1251 TGACGCTAAAGGCTTCCGAATTTTGACAGCACGCTGAATAT..... 1293
329 .AspGlyAlaGlnGlyLysHisIleSerGlyHisArgAsnTyrIleGln 345
1294 .....GATACGAAGCTCGATATTCACGATATTCGCGGCGC 1329
345 LysSerThrLeuAsnGlnAsnIleAsnProGlnGlnLeuLeuAsnGly 361
1330 GGATATCAAGGCTGCTGTTGATGCGAAACGAGG 1371
362 ILeHisSerGlyAlaTyrProValIleSerLysGlyAlaArg 375

```

seq.name: sp_bacteria:Q9X6P3

seq_documentation_block:

```

ID Q9X6P3 PRELIMINARY; PRT; 509 AA.
AC Q9X6P3;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE ADHESIN MAFB.
GN MAFB.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS11;
RA Elckern-Jeager S., Meyer T.F., Fischer E., Maier J., Manning P.A.,
RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF142582; M31039.1;
SQ SEQUENCE 509 AA; 55681 MW; 88B420CFB8269D4 CRC64;

```

alignment_scores:

```

Quality: 450.00 Length: 417
Ratio: 1.875 Gaps: 18
Percent Similarity: 57.554 Percent Identity: 32.374

```

alignment_block:

US-09-303-518D-463 x Q9X6P3 ..

Align seg 1/1 to: Q9X6P3 from: 1 to: 509

```

31 CTGTCCATATCGCAGTGTGCTGCGCGATGCATGCACACGCCTCAGATT 80
16 IleSerLeuLeuGlnIleProIle.....SerHisAlaAsnGlyLe 29
81 GGCAAACGATCCCTTATCCGCGAGTTCTGACCGTCGACGATTCGACAC 130
29 u.....AspAlaArgLeuArgAspMetGlnAlaIleHisIleTyrGln 44
131 CCGAGCGGAATACCACTATTCGCG...AGCAGGGGAGAGCTTGCCNAG 177
44 TGGlyGlyLysTyrHisLeuPheGlyAsnAlaArgGlySerVal..... 58
178 CGCAACGCGCATATCGGATTCGGAACATACAAAGCCATCGATTCGGGCA 227
59 LysAsnArgValGlyAlaValGlnThrPheAspAlaThrAlaValGlyP 75
228 CCGTATG.....ATCAACAGGCGCGCTTGAAGAAATATTCGCGT 268
.....

```



```

75 otleuProlleThrhNlsGluArgThrGlyPheGluGlyIleIleGlyT 92
269 ACATTGTCGGCTTTTCCGATCAGGGGACAAATTCATTCGCGCTTGAC 318
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
92 yGluThrhNlsPheSerGlyNlsGlyIleValNlsIleSerProPheAsp 108
319 AACCATGCGCTCACATTCGATTCGATCAGGAGCGGATGCGCTTGACG 368
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109 AsnNlsAspSerIleSerThrhSerPheSerGlyGlyValAspGlyI 125
369 ATTACGCTTTTACCGCATTCGATTCGATCAGGATCAGCATCCGCGCG 418
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125 yPheThrhValIleGlyIleNlsIleArgThrGlySerGluIleNlsProAla 142
419 ACGGCTATGACGGGCGACAGGGGCGGCTATCCGCTCCCAAGCGCG 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
142 spElyTrAspGlyProGlnGlyGlyIleTyPheProGlnProGlnGlyAla 158
469 AGGATATATACAGTACGACATCAAAAGCGGTGCGCAAAATATCCGCT 518
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
159 ArgAspIleTySerIleThrhNlsIleGlyIleThrhSerThrhIleThrhLys 175
519 CAAC.....CTGACGACACAC 535
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 eAsnThrhValProGlnAlaProPheSerAspArgTrpLeuLysGluAsn 192
536 GCAGCAGCGGACAAAGCGCTTCCGACCGCTTCCACAAAGCGGCGGTATG 585
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
192 IeAlaIleAlaSerGlyPheLeuSerArgAlaAspIleValGlyIleLeu 208
586 CTGACGACAGAGTACGACGAGTTCACAAAGCGGCGGCGGATACAGCC 635
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
209 IleTrpIleAsnAspProAspLysAsnTrpArgAlaAsnArgMetAsp 224
636 CAGAGCTGCACAGATCGGGACATCCCGGCAAGCGCTTCACAGCGACTGCA 685
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 .....AspIleArgGlyIleValGlnGlyAlaValAsn..... 235
686 ATATGTCGCAAAACATCATCGCGCGGACGAGAGAAATTCGCGCGCA... 732
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 .....ProPheLeuThrhGlyPheGlnGlyLeuGlyValGlyAlaIle 249
733 ...GGCGATGCGCTGCAGGCTATACGGAAGGCTCA.....AACAT 770
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 ThrAspSerAlaValSerProValThrhIleAlaIleAlaIleArgIleThrIle 266
771 TGCTGTGATGACAGCGCTTGGCTGCTGCTTCCAGGAAACAAAGATGGCG 820
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 uGlnGlyIleHisAsnLeuGlyAsnLeuSerProGlnAlaGlnLeuAla 283
821 GCATCAGCATTTGCGATATGCG...CAACTCAAAAGACTATGCGCGCA 867
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 IeAlaSerIleLeuGlnAspSerAlaPheAlaValLysAspGlyIleAsn 299
868 GCAGCCATCGCGGATTCGAGAGTCCAAACCCCAATCCCGGCAAGGAT 917
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 SerAla...ArgGlnTrpAlaAspAlaHisProAsn.....I 311
918 AGAAGCGGTCAGCAATATCTTTATGCGACGATCCCATCAAGGAGTGG 967
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
311 eThrhIleThrhAlaGlnIleThrhAlaValAlaValAlaGlyIle... 326
968 GAGTGTCCGGGAAATACGCGCTTGGCGGCGATCAGGACATCTGTGCT 1017
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
327 .....ThrhValTrpGlyGlyLysValGlnLeuAs 337
1018 AAGGCGTGGC...AGATGGCGGCGATGCGCATGGCGAAAGGAATCCG 1064
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 nProThrhLysTrpAspTrp..... 343
1065 CGTGACGACAAATTTTGGCGGATGCGGACATAGCCAAATACCGCTCCCT 1114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
344 .....ValLysAsnThrhGlyTrpGlnLysProAla 353

```

```

1115 AACATCCCGAAATATCCGTTCAAACTTGAGCAGCGCTTACGGCAAGA 1164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 AlaArgPro.....MetGlnThrhValAspGlyGlnMetAlaGlyLys 367
1165 AACATCAGCTCTCCAAACCGTGGCGCGCTCAACGGCAAAATATCAAC 1213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 sAsnLysProProLysProSerThrhGlnGlnHisSerThrhNlsSerAsp 383

seq_name: sp_bacteriap.09JWK8
seq_documentation_block:
ID 09JWK8 PRELIMINARY; PRT; 482 AA.
AC 09JWK8;
DT 01-OCT-2000 (Tremblrel, 15, Created)
DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE ADHESIN.
GN MAFB OR NMA0324.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajadheem M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506(2000).
DR EMBL: AL162752; CAB83629.1; -.
KW complete proteome.
SQ SEQUENCE 482 AA; 52339 MW; 8D3C9EB886ED3DA CRC64;

alignment_scores:
Quality: 414.50 Length: 457
Ratio: 1.692 Gaps: 20
Percent Similarity: 53.611 Percent Identity: 30.416

alignment_block:
US-09-303-518D-463 x 09JWK8
Align seg 1/1 to: 09JWK8 from: 1 to: 482

31 CTGTCCATACCTGCGAGTGTGCTGCCGATGCATGCACAGCGCTCAGATT 80
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
16 ILeSerLeuGlnIleProIle.....SerHisAlaAsnGlyLe 29
81 GGCAAAAGATCCCTTATCCGCGAGGTTCGACCGCTCAGCATTTGCAAC 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29 u.....AspAlaAlaGlnAlaArgAspMetGlnAlaLysHisTrpGly 44
178 CGCAACGGCATATCGGATTTGGGAAACATACAAACCATCAGTTCGGGCA 227
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
59 LysAsnArgValAlaTrpAlaValAlaGlnTrpPheAspAlaTrpAlaValAlaGly 75
228 CCGATG.....ATTCAACAGCGCGCGTGAAGAAATATACGCT 268
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
75 otleuProlleThrhNlsGluArgThrGlyPheGluGlyIleIleGlyT 92
269 ACATTGTCGGCTTTTCCGATCAGGGGACAAATTCATTCGCGCTTGAC 318
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
92 yGluThrhNlsPheSerGlyNlsGlyIleValNlsIleSerProPheAsp 108

```



```

319 AACATGCGCTCACATTCGATTCGACGAGCGCGTAGTCCCGTTGACGG 368
||||| ||| ||| ||||| :|||: |||
109 AsnHisAspSerLysSerThrSerAspPheSerGlyValAspGly 125
||||| :|||: ||| ||| |||||
369 ATTCAGCCTTTACCGCATTCATTTGGAGCATAGCAACACCATCCCGCG 418
||||| :|||: ||| ||| |||||
125 yPheThrValTyrGlnLeuHisArgThrGlySerGlnLeuHisProGlu 142
||||| :|||: ||| ||| |||||
419 ACGGTATGAGCGCGCCACAGCGCGCGCTATCCCGCTCCCAAGCGCG 468
||||| :|||: ||| ||| |||||
142 SPGLTYTAspGlyProGlnGlySerAspTyrProProGlyGlyVal 158
||||| :|||: ||| ||| |||||
469 AGGATATATACAGTACGATCAATAAAGCGTTCGCCCAATATCCGCT 518
||||| :|||: ||| ||| |||||
159 ArgAspIleTyrSerTyrTyrValLysGlyThrSerThrLysThrLysSe 175
||||| :|||: ||| ||| |||||
519 CAACCTGACCGCAACCGCGACCGCGCAACGCGCTTGGCGACCGTTTC 567
||||| :|||: ||| ||| |||||
175 rAsnIleValProArgAlaPro.....PheSerAspArgTrpL 188
||||| :|||: ||| ||| |||||
568 .....CACATGCGCGCGCTATGCTGACGCAAGAGTAGCGAGCGATTC 612
||||| :|||: ||| ||| |||||
188 euLysGluAsnAlaGlyAla.....AlaSerGlyPhe 198
||||| :|||: ||| ||| |||||
613 .....AACGCGCCACCGCATACAGCCCGCA 638
||||| :|||: ||| ||| |||||
199 PheSerArgAlaAspGluAlaGlyLysLeuIleTrpGlnSerAspProAs 215
||||| :|||: ||| ||| |||||
639 GCTGAGACAGATCGGCGCATCGCGCGACGCTTCACGCGACTGCAGATA 688
||||| :|||: ||| ||| |||||
215 nLysAsnTrpTrpAlaAsnArgMetAspAspIleArgGlyIleValGln 232
||||| :|||: ||| ||| |||||
689 TCGTCAAAAC.....ATCATCGCGCGCGAGAGAGAAATTCGCGGCA 732
||||| :|||: ||| ||| |||||
232 LysAlaValAsnProPheLeuMetCylGheGlnGlyAlaGlyIleGlyAla 248
||||| :|||: ||| ||| |||||
733 .....GGCGATCCGTCAGGCGTAAAGGAAAGGCTCAAAATTCGCTG 776
||||| :|||: ||| ||| |||||
249 IleThrAspSerAlaValSerProValThrAspThrAlaAlaGlnGlnTh 265
||||| :|||: ||| ||| |||||
777 CATGACAGCG.....TTGGTCTGCTTCACCGCAAAACAAACATGG 817
||||| :|||: ||| ||| |||||
265 rLeuGlnGlyIleAsnHisLeuGlnAsnLeuSerProGlnAlaGlnLeu 282
||||| :|||: ||| ||| |||||
818 CGCGCATCAACGATTTGGCAGATATGGCG...CAACTCAAGACATATGCC 864
||||| :|||: ||| ||| |||||
282 laAlaIleThrAlaLeuGlnAspSerAlaPheAlaValLysAspGlyIle 298
||||| :|||: ||| ||| |||||
865 GCAGCAGCCATCGCGATTTGGCAGATCAAAACCCCATATGCCGACAGG 914
||||| :|||: ||| ||| |||||
299 AsnSerAla...ArgGlnTrpAlaAspAlaHisProAsnIleThrAlaTh 314
||||| :|||: ||| ||| |||||
915 CAAAGAACCGGTCAGCAATCTTATGGCAGCAGCATCCCAAGAGG 963
||||| :|||: ||| ||| |||||
314 rAlaGlnThrAlaLeuAlaValAlaGlnAlaAlaThrThrValTrpGly 331
||||| :|||: ||| ||| |||||
963 ..... 963
331 LysLysValGlnLeuAsnProThrLysTrpAspTrpValLysAsnThr 347
||||| :|||: ||| ||| |||||
964 .....ATTGAGCTGTCCGCGGAAATATCGGCTTGGCGGATTCAC 1004
||||| :|||: ||| ||| |||||
348 GLTyrLysThrProAlaValArg.....Th 356
||||| :|||: ||| ||| |||||
1005 GGCACATCCTGTCAACGGTGCAGATGGCGGATCGCATTCGCCAAG 1054
||||| :|||: ||| ||| |||||
356 rMetHisThrLeuAspGlyGlnMetAlaGlyLysAsnArgProPro... 371
||||| :|||: ||| ||| |||||
1055 GGAATCCCGGCTCAGCAGCAATTTTCCGATCG..... 1089
||||| :|||: ||| ||| |||||
372 ..LysSerIleThrSerAsnSerLysAlaAspAlaSerThrGlnProSer 387
||||| :|||: ||| ||| |||||

```

```

1090 .....GCATAGCC 1097
|||||
388 LeuGlnAlaGlnLeuIleGlyGlnGlnIleSerSerGlyHisAlaTyrAs 404
|||||
1098 CAATATC.....CCGTCCTTACCATTCCTCCCAATATCCGTCA 1138
||||| :|||: ||| ||| |||||
404 nLysHisValIleArgGlnGlnIleGlnIleThrAspLeuAsnIleAsnSerP 421
||||| :|||: ||| ||| |||||
1139 ACTTGGAGCAGCGTTACGGCAAGAAACATCACTCTCAACCGTGGCG 1188
||||| :|||: ||| ||| |||||
421 rAlaAspPheAlaArgHisIleGlyAlaAsnIleValSerHis..... 434
||||| :|||: ||| ||| |||||
1189 CCGTCAAGCGCAAAATGTC 1209
||||| :|||: ||| ||| |||||
435 ProThrAsnMetLysGlnLeu 441
||||| :|||: ||| ||| |||||

```

```

seq_name: sp_bacteriap:Q9JXD2
seq_documentation_block:
ID Q9JXD2 PRELIMINARY; PRT: 483 AA.
AC Q9JXD2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MARB PROTEIN.
GN NMB2105
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBT_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B.
RX MEDLINE=201755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
RA Cotton M.D., Uitterlinden T.R., Kouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizze M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL; AF002559; AAF62340.1; -.
KW Complete proteome.
SO SEQUENCE 483 AA; 52789 MW; D6B5AEFA8AEBA02 CRC64;

```

```

alignment_scores:
Quality: 390 00 Length: 433
Ratio: 1.618 Gaps: 25
Percent Similarity: 55.658 Percent Identity: 30.716

```

alignment_block:

US-09-303-518D-463 x Q9JXD2 ..

Align seg 1/1 to: Q9JXD2 from: 1 to: 483

```

31 CTGTCCATACATGCGAGTGTGCTGCGCATGCATGCACAGCCCTGAGATT 80
||||| :|||: ||| ||| |||||
16 IleSerLeuLeuGlnIleProIle.....SerHisAlaAsnGlyLe 29
||||| :|||: ||| ||| |||||
81 GGCAGAACATCCCTTTATCCGCGCAGGTTCTGACCCGTCAGATTTGAC 130
||||| :|||: ||| ||| |||||
29 u.....AspAlaArgLeuArgAspAspMetGlnAlaLysHisTyrGlnP 44
||||| :|||: ||| ||| |||||
131 CGAGCGGGAATACCACTATTCGCG...AGCAGGGGAGGCTTGCCNAG 177
||||| :|||: ||| ||| |||||
44 rGlyGlyLysTyrHisLeuPheGlyAsnAlaArgGlySerValLysLys 60
||||| :|||: ||| ||| |||||
178 CGCAAGCGCATATCGATTGGGAAACATACAAAGCCATCAG..... 219
||||| :|||: ||| ||| |||||

```



```
263 TCAGCTACATTTGTCGCTTTTCGATCGAGGCGCAAAATTCATTTGCGCC 312
||||| ||| ||||| ||||| |||||
291 gATgTleHserPro...PheArgSerArgSerArgSerProIleArga 307
||||| ||||| ||||| ||||| |||||
313 TT...CGACAAACATGCTCTACATTCGATCTTGACGAAGCCGTAAGTCC 359
||||| ||||| ||||| ||||| |||||
307 rGHISArgArgProThrHISgluGlyArgArgGlnSerProAlaProSer 323
||||| ||||| ||||| ||||| |||||
360 CGTTGACGATTCAGCCTTTACCGCATTCATGGAGCGATACGAACACC 409
||||| ||||| ||||| ||||| |||||
324 ArgArgArgArgSerProSerPro...Pr 332
||||| ||||| ||||| ||||| |||||
410 ATCCCGCGGAGCTATGACGGCGGCGGCTATCCCGCTCCC 459
||||| ||||| ||||| ||||| |||||
332 oAlaArgArg...ArgSerProSerProAla 343
||||| ||||| ||||| ||||| |||||
460 AAAGGCGGAGGATATATACAGTACGACATAAAGCGCTGGCCGAAA 509
||||| ||||| ||||| ||||| |||||
343 laArgArg...ArgArgSerProSer 350
||||| ||||| ||||| ||||| |||||
510 TATCCGCTCAACCTGACCGACACCG...CAGCAGCG 544
||||| ||||| ||||| ||||| |||||
351 ProProAlaArgArgHISArgSerProThrProProAlaArgGlnArg 367
||||| ||||| ||||| ||||| |||||
545 GACACAGGCTTGCCGACCGCTTCCACAAATCCGCGCTATGTCGACGCA 594
||||| ||||| ||||| ||||| |||||
367 gSer...ProSerProProAlaArgArgHISArgSerPro 380
||||| ||||| ||||| ||||| |||||
595 GAGTACGACGAGATTCAAACGGCGGCGGATACAGCCCGCAGCTGGG 644
||||| ||||| ||||| ||||| |||||
380 roProAlaArgArg...ArgArgSerProSerProPro... 391
||||| ||||| ||||| ||||| |||||
645 CAGATCGGCAATGCCCGCGAAGCCTTCAAGCGCAGCTATATGCTCA 694
||||| ||||| ||||| ||||| |||||
392 ...AlaArgArgArgArgSerProSerProProAlaArgArgArg 406
||||| ||||| ||||| ||||| |||||
695 AAAACATCATCGCGCGCGGAGGAAT...TGTC 726
||||| ||||| ||||| ||||| |||||
406 gSerProSerProLeuTyrArgArgArgArgSerProSerProLeuTyr 423
||||| ||||| ||||| ||||| |||||
727 GGGCAGCGCATGCGCTGACGAGGTATAGCGAGGCTCAAACTGCTCT 776
||||| ||||| ||||| ||||| |||||
423 rGArgAsnArgSerArgSerProLeuAlaArgGlyArgSerAspSer 439
||||| ||||| ||||| ||||| |||||
777 CATGACAGGCTTGCTGCTTCCACCGAAACAGAT...GGCGC 820
||||| ||||| ||||| ||||| |||||
440 ProGlyArgSerProSerProValAlaArgLeuArgAspProThrGly 456
||||| ||||| ||||| ||||| |||||
821 G...CATCAGCATTTGGCAGATATGGCGCA 849
||||| ||||| ||||| ||||| |||||
456 ArgLeuProSerProSerIleGlnArgLeuProSerProProVala 473
||||| ||||| ||||| ||||| |||||
850 CTGAAGACTATGCCGACAGCCATCCGCAATGGGAGTCCAAAACCC 899
||||| ||||| ||||| ||||| |||||
473 laGlnArgLeuProSerProProArgArgArgArgLeuProSerPro 489
||||| ||||| ||||| ||||| |||||
900 CAATGCCGACAAAGCATAGAAC... 923
||||| ||||| ||||| ||||| |||||
490 ProProAlaGlnArgLeuProSerProProArgArgArgArgLeuPr 506
||||| ||||| ||||| ||||| |||||
924 .CGTCGCAATATCTTTATGCGAGCAT...C 951
||||| ||||| ||||| ||||| |||||
506 oSerProMetArgIleGlySerHISAlaAlaAsnHISLeuGlnSerP 523
||||| ||||| ||||| ||||| |||||
952 CCCATTAAGGATTTGGAGCTGTCGGGGAATAACGGCTT... 992
||||| ||||| ||||| ||||| |||||
523 roSerProSerLeuSerProProGlyArgGlyLysValLeuProSer 539
||||| ||||| ||||| ||||| |||||
992 ... 992
||||| ||||| ||||| ||||| |||||
540 ProProValaArgArgArgArgSerLeuThrProAspGlnGlyArgVal 556
||||| ||||| ||||| ||||| |||||
993 ...GGCGGATCATCAGCGCACATCTGTCAAGCGGTGCGAG 1029
||||| ||||| ||||| ||||| |||||
```

```
556 rLeuSerGlnGlyArgHIS_ThrSer...ProSerHISLeuGln 571
||||| ||||| ||||| ||||| |||||
1030 ATGGCGCGCATGCCA...TGCCGAAAGGCAAAATCCGCCGCTCAGCA 1073
||||| ||||| ||||| ||||| |||||
572 AspGlySerMetSerProValaArgGlyArgGlyLysSerProSerSe 588
||||| ||||| ||||| ||||| |||||
1074 CAATTTTGGCGATGC...GCATACGCGCAAAATCCCGTCCCTT 1114
||||| ||||| ||||| ||||| |||||
588 rArgHISGlnLysAlaArgSerProValaArgArgArgSerProThrPro 605
||||| ||||| ||||| ||||| |||||
1115 ACCATTCGCCGAATATCCGTTCAAACTTGACAGCGT...TAC 1155
||||| ||||| ||||| ||||| |||||
605 AlaAsnArgArgSerArgSerSerArgSerAlaSerArgSerProAspArg 621
||||| ||||| ||||| ||||| |||||
1156 GGCAGAAAGAAACATCACCTCTCAACCGCGCGGCTCAAGGCGAANA 1205
||||| ||||| ||||| ||||| |||||
622 ArgArgArgArgSerProSerSerArgSerArgSerArgSerArgSe 638
||||| ||||| ||||| ||||| |||||
1206 TGTCAAACTG... 1216
||||| ||||| ||||| ||||| |||||
638 rProProValaLeuHISArgSerProSerProArgGlyArgLysHISGln 655
||||| ||||| ||||| ||||| |||||
1217 CAGACCAAGCGCACCG... 1233
||||| ||||| ||||| ||||| |||||
655 rGlnArgArgSerProGlyArgLeuSerGlnGlnAspArgValGln 671
||||| ||||| ||||| ||||| |||||
1234 ...AAGCAGCGCTACCGCTTGTGACGTAAGGGTT 1265
||||| ||||| ||||| ||||| |||||
672 AsnSerLysLeuLeuLysArgThrSerVal... 681
||||| ||||| ||||| ||||| |||||
1266 TCCGAATTTTGGAAGCAGCTGAATATATGACAGCTCGATATTCAAG 1315
||||| ||||| ||||| ||||| |||||
682 .ProAspThrAspLysArgLysGlnLeuProGlnLysLeu... 694
||||| ||||| ||||| ||||| |||||
1316 AATTATCGGGGGGCGGTATACCTAAGGCTAAGCCGTGTTGATGCCAAA 1365
||||| ||||| ||||| ||||| |||||
695 ..LeuGlnValGlyArgValGlnHISLys... 704
||||| ||||| ||||| ||||| |||||
1366 CCGACATGGAGGTTGATAGACGCTTAATAATTTAGCACTCGTAGCA 1415
||||| ||||| ||||| ||||| |||||
705 ...GlnGlnArgLysSerAspLysSerGlnLysArgSe 718
||||| ||||| ||||| ||||| |||||
1416 GGTGAGAAAT...GTTACAGCAA 1438
||||| ||||| ||||| ||||| |||||
718 rValHISArgHISISGlySerGlnMetSerProValGlnAsnSerGlnG 735
||||| ||||| ||||| ||||| |||||
1439 CGAAGAAGAGATCGAGTAGTCAATTAAAGCCATGCCAGACAGAGA 1488
||||| ||||| ||||| ||||| |||||
735 LysArgSerArgProValSerSerLysValLysAspSerGlnGlnValG 751
||||| ||||| ||||| ||||| |||||
1489 TGGGAAATTAACAGGCTTAGATTTTATCATCAATTTTAGGTGATAT 1558
||||| ||||| ||||| ||||| |||||
752 LysGlnAspAsnSerAspLeuAspAlaAsnLeu...SerCy 764
||||| ||||| ||||| ||||| |||||
1539 CAATAGAAAGGACAGTAACAGAGGCGCATAGCTA...A 1576
||||| ||||| ||||| ||||| |||||
764 sAspSerLysAspThrIleArg...HISGlnIleLysAspLysAsnA 779
||||| ||||| ||||| ||||| |||||
1577 CCCGTTGATGTAACGGGTATACACAAACCTGGACCTGATTAACAT 1626
||||| ||||| ||||| ||||| |||||
779 rArgGlyAsnLysArgSerSerArgGlnGlnValSerSerAspAspAsn 795
||||| ||||| ||||| ||||| |||||
1627 GGGTTTATCAAGGACAGCTGGAATTTAAAGCTGATGGAAGCTGGGA 1676
||||| ||||| ||||| ||||| |||||
796 GlySerSerAspSerAspValAspAspArgLys...G1 807
||||| ||||| ||||| ||||| |||||
1677 GGTGAAGAAAGAAAGGTGGGAAGTGAAGACCAAG 1713
||||| ||||| ||||| ||||| |||||
807 uAlaLysArgArgArgGlyGlnGlnLysLysThrArg 819
||||| ||||| ||||| ||||| |||||
```

seq_name: sp_plant:092W08


```

644 GlyArgSerArgThrProAlaArgGlyArgSerArgSer 660
776 TCATGCACGGCTGGTCTGCTTCCACCGA.....AACACAGATG 816
660 gThProArgArgGlyArgSerArgSerLeuValArgArgGly 677
817 GCGGCGCAT.....CAACGATTTGGCAGATATGGCGCA..... 848
677 rGSerHisSerArgThrProGlnArgArgGlyArgSerGlySerSer 693
849 .....ACTCAAGACTATGCGGACG...AG 871
694 GlnArgLysAnlySerArgThrSerGlnArgArgSerArgSer 710
872 CCATCCGCA.....T 882
710 rSerProGlnMetLysLysSerArgIleSerSerArgArgSer 727
883 TGGGCACTCCAAAACCCCAATGCGGCACAGCATAGAACCGTCACAA 932
727 euserSerProArgSerLysAlaLysSerArg...LeuSerLeuArgArg 742
933 TATCTTATGGACACCAT.....CCCCAACAAGGATTCAGCTGTCC 976
743 SerLeuSerLysSerProCysProLysGlnLysSerGlnThrPro 759
977 GGGGAAATACGGCTTGGCGGCATCAC.....GGACAT 1011
759 oArgArgSerArgSerLysSerSerGlnProLysAlaLysSerArgThr 775
1012 CCTGTCAAGGGTTCAGATGGCGCGCATTCGCCGAAGGAAGATC 1061
776 ProProArgArgSerArgSerSerSerProProLysGlnLys 792
1062 CGCGGTACGACATTTGGCGATGGGATACGCCAATACCCCTCCC 1111
792 rLysThrProSerArgLysSerHisSerSerSerProHisPro 807
1112 CTATACATTTCCGAATATCCGTTCAACTGGAGCAGCTTACGGCAAA 1161
808 .....LysValLysSerGlyThrProProArgGlnGly... 818
1162 GAAACATCATCCTTCACACCGTCCGCCGTCAACAGCGCAAAATGCAA 1211
819 .....SerIleThrSerProGlnAlaAsnGlnLysValInh 831
1212 ACTGGAGACCAACGCCCGGACAGACAGCGCTTTCAGCGTAAG 1261
831 r.....ProGlnArgArgSerCysPheGlnSerSerP 842
1262 GGTTCGCAATTTTGAAGACAGTGAATATGATACGAAGCTCGATTT 1311
842 roAspProGlnLeuLysSerArgThrPro.....SerArg 853
1312 CAAGATTTATCGGGGGCGGTATACCTAAGGCTTGTGATGATC 1361
854 HisSerCysSerLysSerProProArgValLysSer.....SerThr 868
1362 GAACCCGAGATGGAGTGTATAGAACTTATAAATGACACACTCGTG 1411
868 rProProArgLysLysSerProSerArgSer.....SerSerProG 881
1412 AGCAGGTGAGAAAATGTTCAAGAAACGAGAGAGAGAGTACAGT 1458
881 LysProLysValLysAlaIleIleSerProArgGlnArgSerHisSer 896
seq_name: sp_human:09U040
seq_documentation_block:
ID 09U040 PRELIMINARY; PRT; 1262 AA.
AC 09U040;
DT 01-MAY-2000 (TRENBLREL, 13, Created)
DT 01-MAY-2000 (TRENBLREL, 13, Last sequence update)

```

```

DT 01-DEC-2001 (TRENBLREL, 19, Last annotation update)
DE RNA BINDING PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016087; BAA83713.1;
DR InterPro; IPR002965; P_Rich_extensn.
DR PRINTS; PR01217; PRICHEXTENSN.
FT NON_TER 1
FT NON_TER 1262
SQ SEQUENCE 1262 AA; 138003 MW; 75E76BD73FA14295 CRC64;

alignment_scores:
Quality: 157.00 Length: 617
Ratio: 0.584 Gaps: 30
Percent Similarity: 43.598 Percent Identity: 22.690

alignment_block:
US-09-303-518D-463 x 09U040 ..

Align seg 1/1 to: 09U040 from: 1 to: 1262

11 CCCGAAATATCCCTTATTTGTCATACAGCGCATGCGCGATG 60
||||| |||||::: |||||
154 ProAlaProThrProLeuLeu.....AlaGlnArgHis 164
61 CATGCACAGCCCTCAGATTGGCAACGATCCCTTATCCGCGAG...T 107
:::||||| ||| |||||:::
164 sGlyLysSerProGlnProLeuAlaThrThrProLeuSerGlnPro 181
108 TCTGCAGCGTCAGCATTTCCGAACCCGAGGGGAAATACAC... 148
||| ||| :|||::: |||
181 AlaAsnProProSerGlnAlaSerProThrArgAspArgSerProLys 197
149 .....TATTCGACAGCAGGGGAGCTTGGCCAG 177
||||: |||::: |||||
198 SerProGlnLysLeuProGlnSerSerSerSerGlnSerProProSe 214
178 CGCAAGCGCCATATCGGATTGGAAACATACA..... 209
| :||| :|||::: |||
214 r.ProGlnProThrLysValSerArgHisAlaSerSerProGlnSer 230
210 ..AAGCATCATGTTGGGCA..... 227
|||||::: |||||
221 ProLysProAlaProAlaProGlySerHisArgGlnLysSerSerP 247
228 .....CTGATGATTCACAGCGCGCGCTTGAAGAAATATCGCTACAT 272
::: ||||| :|||::: |||
247 cThrSerLysAsnArgSerHisGlyArgAlaLysArgAspLysSerHis 264
273 .....TGT 275
264 eHisThrProSerArgArgMetGlyArgSerArgSerProAlaThrAla 280
276 CCGCTTTCCGATCAGCGGCGACAAATTCATTC..... 308
|||||::: |||||
281 LysArgGlyArgSerArgSerArgThrProThrLysArgGlyHisSer 297
309 .....GCCCTTGACAAACCATGCTTCATTCGATTCTGAC 345
||||: |||::: |||
297 gSerArgSerProGlnThrProArgArgSerArgSerAlaGlnArgT 314
346 GAACCGCGTAGTCCCGTTGACAGGATTAGCCTTACCGCATCATTTGGA 395
||||| |||||
314 rGSerArg.....SerProGlnArgArgGly 322

```



```
374 sGlySerProGlnProLeuAlaThrThrProLeuSerGlnGluProV 391
    ::::|
108 TCTCCAGCGTCAGCATTTCCGAACCCGACGGAATATCCACC..... 148
    ::|
391 AlaSerProSerGlnLaserProThrArgAspArgSerProProlys 407
    ::|
149 .....TATTCGACGACGGGGGAGCTTCCGACG 177
408 SerProGlnLysLeuProGlnSerSerSerGlnSerSerProProse 424
178 CGCAAGGCCATATCGAGTGGGAACATACA..... 209
424 r.ProGlnProThrLysValSerArgHisAlaSerSerSerProGlnSer 440
210 ...AAGCATAGTTGGGCA..... 227
441 ProLysProAlaProAlaProGlnSerHisArgGlnLysSerSerPr 457
228 ...CCTGATGATTCCACAGCGCGCGCTTGAAGGAATATCGGCTACAT 272
457 OThrSerLysAsnArgSerHisGlyArgAlaLysArgAspLysSerHis 474
273 ..... 275
474 eRhIsthrProSerArgArgMetGlyArgSerArgSerProAlaThrAla 490
276 CCGCTTTCCGATCAGCGGACCAATTCATTC..... 308
491 LysArgGlyArgSerArgSerArgThrProThrLysArgGlnHisSerAr 507
309 .....GCCCTTCGACCAACCATGCTCACATTCGAGTTCTGAC 345
507 gSerArgSerProGlnThrParArgSerArgSerArgSerArgThrPro 524
346 GAAGCGGTAAGTCCCGTTGACGAGATTCAGCTTACCGCATTCATTCGGA 395
524 rSerArg.....SerProGlnArgArgGly 532
396 CGGATACGACACCATCCCGC..... 416
533 ArgSerArgSerProGlnArgProGlnLysThrSerArgSerArgSnhThr 549
417 .....CGACGCG 423
549 nArgArgGlyArgSerArgSerAlaArgArgGlyArgSerHisSerArg 566
424 TATGACGGGCGACAGCGGGGCGTATCCGCTCCCAAGCGCGGACGGA 473
566 ePrAlaThrArgGlyArgSerArgSerArgThrProAlaArgArgGly 582
474 TATATACAGCTACGACATAAAGCGCTTGCCCAAAATATCCGCTCAACG 523
583 .....ArgSerArgSerArgThrProAlaArgArgArgSerAr 595
524 TGACCGCAACCCGACGACCGGACAAAGCGTCCGA...CCGTTTCCAC 570
595 gSerArgThrProThrArgArgArgSerArgSerArgThrProAlaArg 612
571 AATGCGCGCGCTATGCTGACGCAAGAGTAGGAGGAGTTCGAACGCGC 620
612 rGlnLysArgSerArgSerArgThrProAlaArgArgArgSerArgThrArg 628
621 CAACCGATACAGCCCGAGCTGACAGATCGGCAATCGCGCGGAGCGCT 670
629 SerProValArgArgArgSerArgSerArgSerProAlaArgArgSerAr 645
671 TCAA.....CGGCACTGCAGATATCGTCAAAA. 698
645 yArgSerArgSerArgThrProAlaArgArgGlyArgSerArgSerArgT 662
699 ..CATCATCGCGCGCGAGAGAAATTGCGG..... 728
```

```
662 hrProAlaArgArgGlyArgSerArgSerArgThrProAlaArgArgSer 678
729 ...CGACGCGGATCCGTCGACGGGTATTAACGCAAGGCTCAAAACATTCG 775
679 GLyArgSerArgSerArgThrProAlaArgArgGlyArgSerArgSerAr 695
776 TCATGACAGCGCTTGGCTGCTTTCACCGCA.....AACAAAGATG 816
695 gThrProArgArgGlyArgSerArgSerArgSerLeuValArgArgGlyA 712
817 GCGCGCAT.....CAACGATTTGGCAGATATGCGCA..... 848
712 rSerHisSerArgThrProGlnArgArgGlyArgSerGlnLysSerSer 728
849 .....ACTCAAAAGCTATGCCGACG...AG 871
729 GlnArgLysAsnLysSerArgThrSerGlnArgArgSerArgSerAsnSe 745
872 CCATCCCGCA.....T 882
745 rSerProGlnMetLysLysSerArgLysSerSerArgArgSerArgSer 762
883 TGCGCAGTCCAAACCCCAATGCCGACAGCATGAGATGAGCCGTCAGCA 932
762 eUserSerProArgSerLysAlaLysSerArg...LeuSerLeuArgArg 777
933 TATCTTTATGCAACCAT....CCCATCAAAAGGATTGAGCTGTCC 976
778 SerLeuSerGlySerSerProCysProLysGlnLysSerGlnThrProPr 794
977 GGGGAATAATAGGCTTGGCGGCATCAC.....GGCACAT 1011
794 oArgArgSerArgSerGlySerSerGlnProLysAlaLysSerArgThr. 810
1012 CCTGTCCACCGCTGCGAGATGGCGGATCGCATTCGCCGAAGGAATC 1061
811 ProProArgArgSerArgSerSerSerSerProProLysGlnLysSe 827
1062 CGCCGTCACGACAAATTTTGGCGATGCGGATACGCCAAATCCGTCGCC 1111
827 rLysThrProSerArgGlnSerHisSerSerSerSerProHisPro... 842
1112 CTTACCATTCGCCAAATATCCGTTCAAACTTGACAGCAGGTCAGCAAA 1161
843 .....LysValLysSerGlyThrProProArgGlnGly... 853
1162 GAAACATCACCTCTCTCAACCGTCCGCGCTCAACGCGCAAAATGTCAA 1211
854 .....SerLethSerProGlnAlaAsnGlnGlnSerValThr 866
1212 ACTGCGACAGCCACCGCACCGCAACAGAGCGCTTACCGTTTGACGGTAAAG 1261
866 r.....ProGlnArgArgSerCysSnhGlnSerSerP 877
1262 GGTTCGGAATTTGAGAAGCACGTGAATATGATAGCAAGCTGCATATT 1311
877 rOAsPProGlnLeuLysSerArgThrPro.....SerArg 888
1312 CAAGAATTAATCGGGGGGCGTATACCTAAGCTAAGCTGTGTTGAGAGC 1361
889 HisSerCysSerGlySerSerProArgValLysSer.....SerThr 903
1362 GAACCGGATGGGAGGTTGATAGAACCTTAATAAATTCACAACCTCGTG 1411
903 rProProArgGlnSerProSerArgSer.....SerSerProG 916
1412 AGCAGGTGAGAGAAAATGTTACGAGAAAGACAGAGAGGACAGT 1458
916 InProLysValLysAlaLeuLeuSerProArgGlnArgSerHisSer 931
```

seq_name: sp_human:090035
seq_documentation_block:

ID 090035 PRELIMINARY; PRT; 2752 AA.
 AC 090035;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE RNA BINDING PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ohkaki S., Umeki K., Sawada Y.;
 RT "Homo sapiens mRNA for RNA binding protein, complete cds."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016092; BAA83718.1;
 DR InterPro: IPR002965; P_Rich_extensn.
 DR PRINTS: PR01217; PRICHEXTENS.
 SQ SEQUENCE 2752 AA; 259672 MW; 109C6AF181097123 CRC64;

alignment_scores:
 Quality: 157.00 Length: 617
 Ratio: 0.584 Gaps: 30
 Percent Similarity: 43.598 Percent Identity: 22.690

alignment_block:
 US-09-303-518D-463 x 090035 ..

Align seg 1/1 to: 090035 from: 1 to: 2752

```

11 CCGCAAAATATCCCTTATCTGTCATCTGGCACTGCTGCGGATG 60
   ||||| |||||::: |||||
364 ProAlaProThrProLeuLeu.....AlaGluArgH1 374
   |||||::: |||||
61 CATGCACAGCCCTCATGATTTGGCAACGATCCCTTATCCGGCAGG...T 107
   ::|||::: |||||
374 sglYgLySerProGlnProLeuAlaThrThrProLeuSerGlnIuProV 391
   ::|||::: |||||
108 TCCTGCAGCCGACGATTTGGCAACCGCAGGAAATACCCAC... 148
   ::|||::: |||||
391 alasnProSerGlnAlaSerProThrArgaspArgSerProPolys 407
   ::|||::: |||||
149 .....TATTCGACGACGAGGGGAGCTTGCCNAG 177
   |||||::: |||||
408 SerProGlnLysLeuProGlnSerSerSerSerGlnSerProProse 424
   ::|||::: |||||
178 CGCAACGCGCATATCGGATTGGAAACATACA..... 209
   ::|||::: |||||
424 r.ProGlnProThrLysValSerArgHisAlaSerSerProGluSer 440
   ::|||::: |||||
210 ...AAGCATCAGTTGGGCCA..... 227
   |||||::: |||||
441 ProLysProAlaProAlaProGlnLysSerHisArgGlnLysSerSerP 457
   ::|||::: |||||
228 ....CCTGATGATTCAACAGCGCGCTTGAGAAATATCGGCTACAT. 272
   ::|||::: |||||
457 oThSerLysAsnLysSerHisGlyArgAlaLysArgAspLysSerHis 474
   ::|||::: |||||
273 .....TGT 275
   ::|||::: |||||
474 eRhIsthrProSerArgArgMetGlyArgSerArgSerProAlaThrAla 490
   ::|||::: |||||
276 CCGCTTTTCGATCAGCGGCACAAATCCATTC..... 308
   |||||::: |||||
491 LysArgGlyArgSerArgSerArgThrProThrLysArgGlnLysSerAr 507
   ::|||::: |||||
309 .....GCCCTTGACACCAATCCCTCACATTCGATTCGAC 345
   ::|||::: |||||
507 gSerArgSerProGlnTrpArgArgSerArgSerAlaGlnArgTrpAla 524
   ::|||::: |||||
346 GAACCGGTATGCCGTTGACGATTCAGGCTTACCGCATTCATTCGGA 395
   |||||::: |||||

```

```

524 rYSerArg.....SerProGlnArgArgGly 532
   |||||
396 CGATACGACACCATCCGC..... 416
   |||||
533 ArgSerArgSerProGlnArgProGlnYTrpSerArgSerArgAsnThrGl 549
   |||||
417 .....CGACGGC 423
   |||||
549 naArgArgGlyArgSerArgSerAlaArgArgGlyArgSerHisSerArgS 566
   |||||
424 TATGACGGCCACAGGCGCGGCTTATCCGCTCCCAAGGCGCGAGGA 473
   ::|||::: |||||
566 erProAlaThrArgGlyArgSerArgSerArgThrProAlaArgArgGly 582
   ::|||::: |||||
474 TATTTACAGTACACATAAAGCGCTGCCCAAAATATCCGCTCAACC 523
   |||||
583 .....ArgSerArgSerArgThrProAlaArgArgSerAr 595
   ::|||::: |||||
524 TGACCGACACCGCAGCAGCAGCAACGCGCTGCCGA...CGGTTCCAC 570
   ::|||::: |||||
595 gSerArgThrProThrArgArgArgSerArgSerArgThrProAlaArg 612
   ::|||::: |||||
571 AATGCCGCGCTATCTGACGCAAGAGTAGCGGCGGATTCAAAGCGCG 620
   ::|||::: |||||
612 rGlyArgSerArgSerArgThrProAlaArgArgArgSerArgThrArg 628
   ::|||::: |||||
621 CACCGGATACAGCCCGAGCTGCACAGATCGGCATGCCCCGAGAGCT 670
   |||||
629 SerProValArgArgArgSerArgSerArgSerProAlaArgArgSer 645
   ::|||::: |||||
671 TCAA.....CGCACCTGCAGATATGTCAAAA. 698
   ::|||::: |||||
645 ArgSerArgSerArgThrProAlaArgArgGlyArgSerArgSerArgT 662
   ::|||::: |||||
699 ...CATCATCGCGCGCGCAGGAGAAATGTCGG..... 728
   |||||
662 hrProAlaArgArgGlyArgSerArgSerArgThrProAlaArgArgSer 678
   ::|||::: |||||
729 ...CGACGCGATCCGTCGAGGGTATAGCGAAGCTCAACATTCCTG 775
   ::|||::: |||||
679 GlyArgSerArgSerArgThrProAlaArgArgGlyArgSerArgSerAr 695
   ::|||::: |||||
776 TCATGCACAGGCTTGCTGCTCCACCGA.....AACAAGATG 816
   ::|||::: |||||
695 gThrProArgArgGlyArgSerArgSerArgSerLeuValArgArgGly 712
   ::|||::: |||||
817 GCGGCGAT.....CAGCATTTGGCAGATATGGCGCA..... 848
   ::|||::: |||||
712 rYSerHisSerArgThrProGlnArgArgGlyArgSerGlnLysSerSer 728
   ::|||::: |||||
849 .....ACTCAAGACTATGCCGAC...AG 871
   ::|||::: |||||
729 GluArgLysAsnLysSerArgThrSerGlnArgArgSerArgSerAsn 745
   ::|||::: |||||
872 CCATCCGGGA.....T 882
   |||||
745 rSerProGlnMetLysLysSerArgLysSerSerArgArgSerArgSerL 762
   ::|||::: |||||
883 TGGCAGTCCAAACCCCAATCCGCGCAGCAGCATAGAACCGCTCAGCA 932
   ::|||::: |||||
762 eUerSerProArgSerLysAlaLysSerArg...LeuSerLeuArgArg 777
   ::|||::: |||||
933 TATCTTTATGACGACAT.....CCCATCAAGAGGATGGAGTGTCC 976
   ::|||::: |||||
778 SerLeuSerGlySerSerProCysProLysGlnLysSerGlnThrPro 794
   ::|||::: |||||
977 GGGGAAATACGCTTGCGCGCATCAC.....GGCAGAT 1011
   ::|||::: |||||
794 oArgArgSerArgSerGlySerSerGlnProLysAlaLysSerArgThr 810
   ::|||::: |||||
1012 CCTGTCAACGGTCCAGATGGCGCGCATCGATTCGCGAAAGGAATTC 1061
   ::|||::: |||||
811 ProProArgArgSerArgSerSerSerSerProProPolysGlnLysSe 827
   ::|||::: |||||

```



```

1062 CGCCGTCACGCAATTTGGCGATCGCGCATCCGCAATACCGGTC 1111
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
827 rlysthrProserArgInserHisSerSerSerProHisPro.... 842
1112 CTTACCATTCGCCGAATATCCGTTCAACTTGAGACGCGTTAGCGCAA 1161
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|
843 .....LysValLysSerGlyThrProProArgGlnGly... 853
1162 GAAACATACCTCTCAACCGCGCGCGCAAAAGCGCAAAATGTCAA 1211
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
854 .....SerIleThrSerProGlnAlaSerGlnGlnSerValTh 866
1212 ACTGGCAGACCAACGCCACCCGGAAGACAGGCGGTACCGTTGACGTAAG 1261
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|
866 r.....ProGlnArgArgSerCysPheGlnSerP 877
1262 GGTTCGCAATTTTGAGACGACGCTGAATATGTAGTACGACGTCGATTT 1311
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|
877 rAspProGlnLeuLysSerArgThrPro.....SerArg 888
1312 CAAGATATCGGGGGGGGCTATACCTAAGGCTAAGCCTGTGTGATGC 1361
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
889 HisSerCysSerGlySerSerProProArgValLysSer.....SerTh 903
1362 GAACCCGATGCGAGGTTGATAGCAAGCTTAATTAATGACAACTGCTG 1411
      : |||||:::|:::|:::|:::|:::|:::|:::|:::|
903 rProProArgGlnSerProSerArgSer.....SerSerProG 916
1412 AGCAGGTGGAGAAATGTTTCAGAAACGAGACAGAGAGTCAAGCT 1458
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|
916 InProLysValLysAlaIleLeuSerProArgGlnArgSerHisSer 931
seq_name: sp_human:09UP82

```

```

seq_documentation_block:
ID 09UP82; PRELIMINARY; PRT; 944 AA.
AC 09UP82;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MUCIN 4 (FRAGMENT).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9077625; PubMed=9858486;
RA Gipson I.K., Spurr-Michaud S., Moccia R., Zhan Q., Toribara N.,
RA Ho S.B., Gariguilo A.R., Hill J.A. III;
RT "MUC4 and MUC5B transcripts are the prevalent mucin messenger
RT ribonucleic acids of the human endocervix.",
RL Biol. Reprod. 60:58-64(1999).
DR EMBL; AF058803; AAC34750.1; -.
FT NON_TER 1
FT NON_TER 944
SQ SEQUENCE 944 AA; 96314 MW; 5887CE9ECS0855B5 CRC64;

```

```

alignment_scores:
  Quality: 155.50      Length: 500
  Ratio: 0.685        Gaps: 20
  Percent Similarity: 45.400  Percent Identity: 21.800

```

alignment_block:

US-09-303-518D-463 x 09UP82

Align seg 1/1 to: 09UP82 from: 1 to: 944

```

9 TTCCCGCAAAATATCCCTTATCTGTGCATACT..... 41
  ::::|:::|:::|:::|:::|:::|:::|:::|:::|
  ::| |||||

```

```

407 TrPProSerSerPheSerSerLysGlyHisThThThTrPserGlnThrG 423
42 .....GGCAGTGTGCTGCCATGCATGCACACGCCCTCAGATTGGCAAC 87
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|
423 uLeuProSerThrSerThrGlyAlaIleThrArgLeuValThGlyAsn 440
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|
88 GATCCCTTTTCCGCGCAGGTTCTCGACCGTC..... 118
      ||| |||||:::|:::|:::|:::|:::|:::|
440 roserThrGlyThrAlaGlyThrIleProArgValProSerLysValSer 456
119 AGCATTTTCGAACCCGAGGAAATACCACTATTCGGCAGCGAGGGGAG 168
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
457 AlaIleGlyGlnProGlyGlnProThrThrThThThSerHisSerThr 473
      |||:::|:::|:::|:::|:::|:::|:::|:::|
169 CTTGCCNAGCGCAGG.....GCCATATCGATTTGGGAAACATACA 209
      |||:::|:::|:::|:::|:::|:::|:::|:::|
473 rLeuProLysThrThrGlyAlaGlyAlaGlnThrGlnThrGlnGlnG 490
210 AAGCCATCGAGTTGGGCGCCACTGATGTATCAACAGCGCGCGTTGAAGAA 259
      :::|:::|:::|:::|:::|:::|:::|:::|:::|
490 hngGlyThrThrGly..... 494
260 ATATCGGCTACATTTCTCCGCTTTTCGATCAGCGGCAAAATTCATTCG 309
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
495 .....GluAlaLeuLeuSerSerProSerGlySerValThThrGlnMetIle 510
310 CCCTTGACACACCATTCCTACATTCGATTCGACGAG..... 349
      : |||||:::|:::|:::|:::|:::|:::|:::|:::|
510 sThrAlaThrSerProSerSerSerProMetLeuAspArgHisThrSerG 527
350 .....CCGGTAGTCCGCTGACGGATTCAGCTTACCCCATTCAT 391
527 InGlnIleThrThrAlaProSerThrAsnHisSerThrIleHisSerThr 543
392 GGGACGGATACGACACATCCCGCGGAGGCTATGACGGCGGACAGGGGC 441
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
544 SerThrSerProGlnGlnSerProAlaValSerGlnArgGlyHisThrG 560
442 GCGCGGTATCCCGCTCCCAAGGCGCGAGGATATACAGCTACAGACAT 491
      |||:::|:::|:::|:::|:::|:::|:::|:::|
560 nAlaProGlnThrThrGlnGlnSerGln.....ThrIle 572
492 AAAGGCGGTGGCCAAATATCCGCTCAAC..... 523
572 rGSerValSerProMetThrAspThrLysThrValThThThProGlySer 588
524 .....TGACCGAACCCGACGACCGGACGACGAG 552
589 SerPheThrAlaSerGlyHisSerProSerGlnIleValProGlnAspAl 605
553 CTTGCCGACCGTTTCCACAAATGCCGCGCTTACTGACGCAAGAGTAGG 602
      : |||||:::|:::|:::|:::|:::|:::|:::|
605 a...ProThrIleSer.....A 610
603 CGACGATTTCAACGCGCCCGATACAGCCCGGAGCTGACAGATCGG 652
      ||:::|:::|:::|:::|:::|:::|:::|:::|
610 lAlaIleThrPheAlaProAlaProThrGlyAspGlyHisThThThGln 626
653 GCATGCGCGCGAAGCCTTCAAGGCGACATGAGATATGTCGCAAAACATC 702
      |||:::|:::|:::|:::|:::|:::|:::|:::|
627 Ala.....ProThrThrAlaLeuGlnAlaIleProSerSerH 639
703 ATCGCGCGCGCAGAGAAATGTGTGGCGGACGCGATCCGCGCAGGGTAT 752
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
639 sAspAlaThrLeuGlyProSerGlyGlyThrSerLeu..... 651
753 AAGCGAGGCTCAAAACATTTGCTGCATGACGAGGCTTGCTGCTTTTCA 802
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
652 ..SerLysThrGlyAlaLeuThrLeuAlaSerValSerThrPro 667
803 .....CCGAAACACAGATGGCGCGCATCAACGATTTGGCAGATATGCG 846
      ||:::|:::|:::|:::|:::|:::|:::|:::|
668 GlyLysProGlnGlyGlnThrThrSerAla..... 677

```

```

847 CAATCAAGAACTATCCCGGACGGCATTCCGGTAT..... 883
678 .....||||| ||||| :::::|||:::
678 SerAlasertHrSerProaspHrAlaIalMetHrIstHrISG 694
911 AAGCA..... GGCAGTCACAACCATAATGCCGCAC 910
694 InalaslaserThrglualaserSerglyInthrInThrSergluProala 710
911 AAGCA.....TAGAAGCCGTC 927
711 SerSerglySerArghTrHrThrSerAlaglYThAlatHrProSerSerge 727
928 AGCAATATCTTTATGGCACCGCATCCCCATCAAGAAGATTGAAGTCGCG 977
727 rglyAlaserSerglyThrInhrProsergylSergInglyIlesterInserg 744
978 GGGAAAATACGGCTGGCGGGCATCA.....CGGCACATCCGG 1015
744 lgluthrThrArghPheSerSerAsnProSerArghspSerHsthThr 760
1016 TCACAGCGGTCCGAGATGGCGCGATCCGATCCGACGGAAGGAATCCGCC 1065
761 GlaserThrThrgluenuleuserAlaserHisglYlalalle 777
1066 GTACGACACAAATTTGCCGATCGGCATACGCCCAAATACCCGCCCCCTTA 1115
777 oValserThrglyMeAlaserSerileValProglYThrPhenIsProt 794
1116 CCATTCGCCAATATCCGTT.....CAAACTTGACAGCCGTT 1153
794 hrleuserglualaserThralaglarProthrInglyInserSerp 810
1154 ACGCCAAGAAACATCACCTCTCAACCCGCGCGCGTAACG.... 1199
811 ThrserProslaserProglInlu.ThralaIalIeserArghMeta 827
1200 .....CAAAAATGTCAAACTGCGAGA..... 1220
827 laglnThrInArghTrArghTrHrserArghSerglsapHrIleSerleu 843
1221 ..... CCAACGCCGCCC 1232
844 AlaserGlinalaThraspthrPheSerThryAlProthrProthro 859

seq_name: sp_human:076037

seq_documentation_block:
ID 076037 PRELIMINARY; PRT; 1045 AA.
AC 076037;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MUCIN PRECURSOR (FRAGMENT).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON MICOSA;
RA MEDLINE=98285542; PubMed=9620877;
RA Noller S., Montaux N., Maury J.P., Petitprez D., Degand P., Laine A.,
RA "Porcheret N., Albert J.P.";
RT "Human Mucin Gene MUC4: Organization of its 5' Region and of its
RT Polymorphic Central Tandem Repeat Array.";
RN Biochem. J. 332:739-748(1998).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON MICOSA;
RA MEDLINE=20381033; PubMed=10920259;
RA Choudhury A., Montaux N., Wimpenny J.P., Hollingsworth M.A.,
```

RA	Albert J.P.,	Matra S.K.;	
RT	"Human MUC4	cDNA and its	variants in pancreatic carcinoma."
RL	J. Biochem.	128:233-243(2000).	
DR	EMBL: AF000281;	CA03985.1;	..
FW	Signal.		
KT	Signal.	1	27
FT	CHAIN	28	>1045
FT	NON-TER	1045	1045
Q0	SEQUENCE	1045 AA:	106737 MW; 26112DF28E5D60A0 CRC64;

alignment_scores:		
Quality:	155.50	Length: 500
Ratio:	0.685	Gaps: 20
Percent Similarity:	45.400	Percent Identity: 21.800

alignment_block:

Align seg 1/1 to: 076037 from: 1 to: 1045

9	TTCCCGCAAAATATCCCTATTCTGCTCAACT	41
493	TRPProSerSerPheSerSerGlyIysThrThrTrpSerGlnThrL	509	
42GACATGCGCTGCGCATGATCCACACGCGCTCAGATTGGCAAC	87	
509	ULeUProSerThrSerThrGlyAlaAlaThrAlaGlyLeuValThrGlyAsnP	526	
88	GATCCCTTATCCGGCAGGTTCTGACCGTTC	118	
526	roSerThrGlyAlaAlaGlyThrIleProAlaGlyAlProSerIysValSer	542	
119	AGATTTGCAACCGGACGGGAAATACCAACCGTATTCGGCAGGGGGAG	168	
543	AlaIleGlyIleProGlyIleProThrThrIysSerSerIisSerThr	559	
169	CTTGCACACCGGACAGC	209	
559	rIleuProIysThrThrGlyAlaGlyAlaGlnThrGlnTrpThrGlnIuT	576	
210	AAGCATCAATTGGGCGACCTGATGATTTCAACAGGCGCGTTGAAGAA	259	
576	hrGlyThrThrGly	580	
260	ATATGCGGCAATATGCGCGTTTTCGATCGGCGCAAAATTCATTCG	309	
581GlnAlaLeuLeuSerProSerThrIysSerValThrGlnMetIleu	596	
310	CCCTGCACACCACTGCTCACTTCCGATTCTGACAGC	349	
596	StrAlaIleThrSerProSerSerSerProMetLeuAspArgHisThrSerG	613	
350CCGATAGTCCCGTTGACGGATTCAGCCGTTTACCGATTCATT	391	
613	InglnIleThrThrAlaProSerThrAsnHisSerThrIleHisSerThr	629	
392	GGGACGATCCGAACACATCCCGCGGACGGCTATGACGGCGACAGGC	441	
630	SetThrSerProGlnIuSerProAlaValSerGlnArgGlyHisThrGln	646	
442	GGCGGCTATCCGCTCCCAAGACGGCGAGGATATATGACAGTACGAT	491	
646	nalAlaProGlnThrThrGlnIuSerGln	658	
492	AAAAGCGGTGCCCAAAATATCCGCGCTCAAC	523	
658	rgSerValSerProMetThrAspThrIysThrValThrThrProGlySer	674	
524TGACCGCAACCGGACGACGACGACGACG	552	
675	serPheThrIaSerGlyHisSerProSerGlnIleValProGlnAspAl	691	

```

553 CTTCGGACGCGTTTCACATGCGGCGCATGCTGACGCAAGCATAGG 602
691 : |||||:|||||
696 A...Prothriliser.....A 696
603 CGAGCGATCAAGCGCGCCACCCATACAGCCCGAGCTGACAGATCGG 652
696 : |||||:|||||
696 laaIatThrPheAlaProAlaProThrGlyAspGlyHisThrGln 712
653 GCAATGCGCGCGAGCCCTCAAGCGCGATGATGATCGCAAAACATC 702
713 Ala.....Prothriliser.....A 713
703 ATCGCGCGCGAGGAGAAATGTCGCGCGAGCGATGCGCGAGGAT 752
725 : |||||:|||||
725 sasPlatThrLeuGlyProSerGlyGlyThrLeu..... 737
753 AAGGAGGCTCAACATGCTGTCATGACAGCGCTGGCTGCTTCCA 802
738 : |||||:|||||
738 SerLysThrGlyAlaLeuThrLeuAlaAsnSerValAlaSerThrPro 753
803 : |||||:|||||
754 GlyLysProGlyGlyGlyThrSerAla..... 763
847 CAATCAAGACTATGCGCGAGCGATCGCGAT..... 883
764 SerAlaSerThrSerProAspThrAlaAlaMetThrHisG 780
884 : |||||:|||||
780 lnaIaGlySerThrGlyAlaSerGlyGlyThrGlyThrSerGlyProAla 796
911 AAGCA.....TAGAAGCGCTC 927
797 SerSerGlySerAlaGlyThrSerAlaGlyThrAlaThrProSerGly 813
928 AGCAATCTTTATGCGCGATCGCGATCAAGGATGAGATGCGCG 977
813 : |||||:|||||
813 rGlyAlaSerGlyThrThrProSerGlySerGlyGlyThrSerG 830
978 GGGAAATAGCGCTGGCGCGATCA.....CGGCAATCGCG 1015
830 : |||||:|||||
830 lylGlyThrThrArgPheSerSerAspProSerAlaGlySerHisThr 846
1016 TCAGCGGCTGCAATGCGCGCGATCGCGATCGCGAAAGCAATCGCG 1065
847 GlnSerThrThrGlyLeuSerAlaSerAlaSerHisGlyAlaLeu 863
1066 GTGAGGCAATTTGCGCGATCGCGATCGCGAAATACCGCTTCA 1115
863 : |||||:|||||
863 oValSerThrGlyMetAlaSerSerLeuValProGlyThrPheHisPro 880
1116 CCATTCGCGAAATATCGCT.....CAAACTGAGAGCGCTT 1153
880 : |||||:|||||
880 hLeuSerGlyAlaSerThrAlaGlyArgProThrGlyGlnSerSerPro 896
1154 ACGGCAAAAGCAATCAGCTCTCAACGCTGCGCGCTCAAGCG... 1199
897 : |||||:|||||
897 ThrSerProSerAlaSerProGlnGlyThrAlaAlaIleSerArgMeta 913
1200 : |||||:|||||
1200 : |||||:|||||
1220 : |||||:|||||
913 laGlnThrGlnArgThrArgThrSerArgGlySerAspThrIleSerLeu 929
1221 : |||||:|||||
930 AlaSerGlnAlaThrAspThrPheSerThrValProProThrProPro 945
seq_name: sp_human:Q9H4D6
seq_documentation_block:
ID Q9H4D6 PRELIMINARY; PRT; 1107 AA.
AC Q9H4D6;

```

```

DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE MUC4 PROTEIN SPLICED VARIANT SV11.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Choudhury A., Moniaux N., Ringel J., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ277505; CAC14585.1;
SQ SEQUENCE 1107 AA; 113576 MW; 2669ADB5D5A3721 CRC64;

alignment_scores:
Quality: 155.50 Length: 517
Ratio: 0.713 Gaps: 21
Percent similarity: 42.166 Percent identity: 22.631

alignment_block:
us-09-303-518d-463 x Q9H4D6 ..

Align seg 1/1 to: Q9H4D6 from: 1 to: 1107

9 TTCCGCAAAATATCCCTATCTGCTCACT..... 41
493 TrpProSerSerPheSerSerLysGlyHisThrTrpSerGlnThrGly 509
42 : |||||:|||||
42 : |||||:|||||
509 uLeuProSerThrSerThrGlyAlaAlaThrArgLeuValThrGlyAsp 526
88 GATCCCTTATCCGCGAGTCTCGACCGTC..... 118
526 roSerThrArgAlaAlaGlyThrIleProAlaGlyAlaProSerLysValSer 542
119 AGCATTTGCAACCCGCGGAAATACACATTTGCGCGAGCGGAG 168
543 : |||||:|||||
543 AlaIleGlyGlyProGlyGlyProThrThrLysSerSerHisSerThr 559
169 CTTCGCGCGCGCAAG.....GCCATATCGGATTCG..... 199
559 : |||||:|||||
559 rLeuProLysThrThrGlyAlaGlyAlaGlnThrGlnThrGlnGly 576
200 : |||||:|||||
576 hngLysThrGlyAlaLeuLeuSerSerProSerThrSerValIle 592
209 : |||||:|||||
209 AAAGCATCAGTTGGCGCACCTGATGATTAACAGCGCGCTTG...AA 255
593 GlnMetIleLysThrAlaThrSerProSerSerSerProMetLeuAspArg 609
256 GCAATATCGCGCTACATTTGCGCTTTTCGATACGCGGCAAAATCCA 305
609 : |||||:|||||
609 gHisThrSerGlnGlnIleThrThrAlaProSerThrAsnHisSerThr 626
306 TTGCGCGCTTGCAGACATGCT...CACATTCGATTCG..... 343
626 : |||||:|||||
626 lHisSerThrSerThrSerProGlnLysSerProAlaValSerGlnArg 642
344 : |||||:|||||
643 GlyHisThrArgAlaProGlnThrThrGlnLysSerGlnThrThrArg 659
352 GGTAGTCCCGTTGACGATTCAGCTTACCGCATTCATGGAGGAGATA 401
659 : |||||:|||||
659 rValSerProMetThrAsp.....ThrLysT 668
402 CGAACACCATCCCGCGAGCGCTATGACG.....GCCACAGGCGCGCG 445

```

```

668 hVvalThrThrProGlySerThrPheThrAlaSerGlyHisSerProSer 684
446 GCTATCCCGCTCCCAAGGCGGAGGATATATACAGCTACGACATATAAA 495
685 GluIleValProGlnAspAla..... 691
496 GCGCTGGCCCAAAATATGCGCTCACTGACCGCAACCGCAGACCGG 545
692 .....ProThrIleSerAlaIleThrPheAlaProAlaProThrG 706
546 ACAAGCGCTGGCGACCGCTTCCACATCGCGCTTATGCTGACGCAAG 595
706 LysGlyHisThrThrGlnAlaProThrThrAlaLeu..... 718
596 GAGTAGCGCAGCATTCAAACGCGCCACCGATACGCGCCGAGCTGAC 645
719 .....GlnAlaIleProSerSerH 725
646 AGATCGGCGAATGCCCGCAACCTTCACAGCGCCTGACGATATGCTGA 695
725 saspaIatThrLeuGlyProSerGlyThrSerLeu.....SerL 739
696 AACCATCATCGCGCGCAGAGAAATGTGCGCGCAGCGCATGCCGTGC 745
739 ysthrGly..... 741
746 AGGGTATTAAGCGAAGGCTCAACATGCTGTATGACGCGCTGGGTCTG 795
742 .....AlaLeuThrIleuAlaAsnSerValValSerThrPr 753
796 CTTTCACCGCAAAACAAGATGGCGCATCAACGATTTGGCAGATATGAC 845
753 oglyGlyProGluGlyInThrPheSerAla..... 763
846 GCAACTCAAAAGACTATGCCGACAGCAGCATCCGCGATT..... 883
764 ..SerIaSerThrSerProAspThrAlaAlaIleThrHisThrHis 779
884 .....GGGCGATCCCAAAACCCCAATGCCGCA 909
780 GlnAlaGlnSerThrGlnAlaSerGlyInThrGlnThrSerGluProAl 796
910 CAAGGCA.....TAGAAGCCGT 926
796 aserSerGlySerArgThrThrSerAlaGlyThrAlaThrProSerSerS 813
927 CAGCAATATCTTATGAGCAGCATCCCATCAAGGAGATGAGCTGTCC 976
813 erGlyAlaSerGlyThrThrProSerGlySerGluGlyIleSerThrSer 829
977 GGGGAAATACGGCTGGGCGGATCA.....CGGACATCT 1014
830 GlyGluThrThrArgPheSerSerAsnProSerArgAspSerHisThrH 846
1015 GTCAAGCGCTCGCAGATGGCGCATCGCATTCGCCAAAGGCAATCCGC 1064
846 rGlnSerThrThrGluLeuLeuSerAlaSerHisGlyAlaIleP 863
1065 CGTCAGCGCAATTTTCCGATCGGATACGCCAATAACCGCTCCCTT 1114
863 rovalSerThrGlyMetAlaSerSerIleValProGlyThrPheHisPro 879
1115 ACCTTCGCCGAATATCGCTT.....CAAACTTGAGCGAGCT 1152
880 ThrLeuSerGluAlaSerThrAlaGlyArgProThrGlyGlnSerSerP 896
1153 TACGGCAAAAGAAACATCACCTCTCAACCGCTCCGCGCTCAACG... 1199
896 ofThrSerProSerAlaSerProGlnGlu..ThrAlaAlaIleSerArgMet 912
1200 .....CAAAATGTCAAACTGGCAGA..... 1220

```

```

913 AlaGlnThrGlnArgThrGlyThrSerArgGlySerAspThrIleSerLe 929
1221 .....CCACGCGCACCC 1232
929 uAlaSerGlnAlaThrAspThrPheSerThrValProThrProPro 945
seq_name: sp_human:Q96ZM2

```

```

seq_documentation block:
ID Q96ZM2 PRELIMINARY; PRt: 1157 AA.

```

```

AC Q96ZM2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE MUC4 PROTEIN SPLICED VARIANT SV13 (MUC4 PROTEIN SPLICED VARIANT
DE SV12).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREATIC, TUDMOUR.
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ400850; CAC14135.1; -
DR EMBL: AJ400849; CAC14134.1; -
SO SEQUENCE 1157 AA; 118925 MW; F46F039320969A9C CRC64;

```

```

alignment_scores:
Quality: 155.50 Length: 517
Ratio: 0.713 Gaps: 21
Percent Similarity: 42.166 Percent Identity: 22.631

```

```

alignment_block:
US-09-303-518D-463 x Q96ZM2 ..

```

```

Align seg 1/1 to: Q96ZM2 from: 1 to: 1157

```

```

9 TTCGCGCAAAATATCCCTATCTGTCCATCT..... 41
493 TrpProSerSerPheSerSerGlyHisThrThrTrpSerGlnThrG 509
42 .....GGCAGTGGCTGGCGCATGCATCAGCAGCGCTCAGATTGGCAAC 87
509 uLeuProSerThrSerThrGlyAlaIleThrArgLeuValThrGlyAsn 526
88 GATCCCTTTATCGCGCAGGTTCTGACCGTC..... 118
526 roserThrArgAlaIleGlyThrIleProArgValProSerIleValSer 542
119 AGCATTTTGAACCGCAGCGGAAATACCATTCGCGAGCAGGAGGAG 168
543 AlaIleGlyGluProGlyGluProThrThrTySerSerHisThrH 559
169 CTGGCCNAGCGCAAG.....GCCATATCGGATTGG..... 199
559 rIleuProIysThrThrGlyAlaGlyAlaGlnThrGlnThrPheGlnIu 576
200 .....GAACATAC 208
576 hrGlyThrThrGlyGluAlaLeuLeuSerSerProSerTyrSerValIle 592
209 AAAGCATCAGTGGGCGCCGATGATTCACAGAGCGCGCTTG...AA 255
593 GlnMetIleIysThrAlaThrSerProSerSerSerPrometLeuAspAr 609
256 GGAATATCGGCTACATTTGCGCTTTTCGATCAGCGGCAAAATTCGA 305
609 gHisThrSerGlnGlnIleThrThrAlaProSerThrAsnHisSerThrI 626

```



```

119 AGATTTCGACCCGACGGGAAATACACCTATTCGGACGACGAGGGGAG 168
      ::::::::::: ||| ||||| ||||| :::::::::::
543 AAlleGIleGIleProGIleProGIleProGIleProGIleProGIle 559
      ::::::::::: ||| ||||| ||||| :::::::::::
169 CTTGCCNACGCAACG.....GCCATATCGGATGG..... 199
      ::::::::::: ||| ||||| ||||| :::::::::::
559 rleuprGIleProGIleProGIleProGIleProGIleProGIle 576
      ::::::::::: ||| ||||| ||||| :::::::::::
200 .....GAAACATAC 208
      ::::::::::: ||| ||||| ||||| :::::::::::
576 hrcGIleProGIleProGIleProGIleProGIleProGIleProGIle 592
      ::::::::::: ||| ||||| ||||| :::::::::::
209 AAAGCATCATGTTGGCCACCTGATATTCACAGGCGCGCTG...AA 255
      ::::::::::: ||| ||||| ||||| :::::::::::
593 GlMetIleGIleProGIleProGIleProGIleProGIleProGIle 609
      ::::::::::: ||| ||||| ||||| :::::::::::
609 ghlstIleProGIleProGIleProGIleProGIleProGIleProGIle 626
      ::::::::::: ||| ||||| ||||| :::::::::::
306 TTGCGCCCTGACACACATGCT...CACATTCGATTCG..... 343
      ::::::::::: ||| ||||| ||||| :::::::::::
626 leHISerIleProGIleProGIleProGIleProGIleProGIleProGIle 642
      ::::::::::: ||| ||||| ||||| :::::::::::
344 .....ACGAGGC 351
      ::::::::::: ||| ||||| ||||| :::::::::::
643 GlnHISerIleProGIleProGIleProGIleProGIleProGIleProGIle 659
      ::::::::::: ||| ||||| ||||| :::::::::::
352 GGTAGTCCCTGACGATTCACGCTTACCGCATTCGACGAGGATA 401
      ::::::::::: ||| ||||| ||||| :::::::::::
659 rleuprGIleProGIleProGIleProGIleProGIleProGIleProGIle 668
      ::::::::::: ||| ||||| ||||| :::::::::::
402 CGAACACCATCCCGCGACGCTATGAC.....GCCACAGGGGGGG 445
      ::::::::::: ||| ||||| ||||| :::::::::::
668 hrcGIleProGIleProGIleProGIleProGIleProGIleProGIle 684
      ::::::::::: ||| ||||| ||||| :::::::::::
446 GCTATCCCGCTCCCAAGCGCGGAGATATACGATACGACATATAA 495
      ::::::::::: ||| ||||| ||||| :::::::::::
685 GlIleProGIleProGIleProGIleProGIleProGIleProGIle 691
      ::::::::::: ||| ||||| ||||| :::::::::::
496 GCGCTTCCCAAAATATCGCTCAACCTGACCGACACGCGACCG 545
      ::::::::::: ||| ||||| ||||| :::::::::::
692 .....ProGIleProGIleProGIleProGIleProGIleProGIle 706
      ::::::::::: ||| ||||| ||||| :::::::::::
546 ACACGCGCTTCCGACGCTTCCACATGCGCGCGCTATGACGACG 595
      ::::::::::: ||| ||||| ||||| :::::::::::
706 lYasnGIleProGIleProGIleProGIleProGIleProGIleProGIle 718
      ::::::::::: ||| ||||| ||||| :::::::::::
596 GAGTAGCGACGATTCACAGCGCGACCGATACAGCGCGCGCTGAC 645
      ::::::::::: ||| ||||| ||||| :::::::::::
719 .....GlIleProGIleProGIleProGIleProGIleProGIleProGIle 725
      ::::::::::: ||| ||||| ||||| :::::::::::
646 AGATCGGCAATCGCGCGACGCTTCAACGCTGACGACGCTGACG 695
      ::::::::::: ||| ||||| ||||| :::::::::::
725 sAsPAlaIleProGIleProGIleProGIleProGIleProGIleProGIle 739
      ::::::::::: ||| ||||| ||||| :::::::::::
696 AAAPATCATCGCGCGCGAGGAGAAATGTCGCGCGAGCGGATCGCTG 745
      ::::::::::: ||| ||||| ||||| :::::::::::
739 ystIleProGIleProGIleProGIleProGIleProGIleProGIle 741
      ::::::::::: ||| ||||| ||||| :::::::::::
746 AGGATATACGAGGCTCAAAATGCTGCTGACGCGCTGCGCTG 795
      ::::::::::: ||| ||||| ||||| :::::::::::
742 .....AlaIleProGIleProGIleProGIleProGIleProGIleProGIle 753
      ::::::::::: ||| ||||| ||||| :::::::::::
796 CTTTCACCGGAAAGATGCGCGCATACAGATTCGACGATATGACG 845
      ::::::::::: ||| ||||| ||||| :::::::::::
753 oGIleProGIleProGIleProGIleProGIleProGIleProGIle 763
      ::::::::::: ||| ||||| ||||| :::::::::::
846 GCAACTCAAAAGATGCGCGAGGACGATCGCGCT..... 883
      ::::::::::: ||| ||||| ||||| :::::::::::
764 ..SerAlaIleProGIleProGIleProGIleProGIleProGIleProGIle 779
      ::::::::::: ||| ||||| ||||| :::::::::::

```

```

884 .....GGCAGCTCCAAACCCCAATGCCGCA 909
      ::::::::::: ||| ||||| ||||| :::::::::::
780 GlIleProGIleProGIleProGIleProGIleProGIleProGIleProGIle 796
      ::::::::::: ||| ||||| ||||| :::::::::::
910 CAAGGCA.....TAGAGCGCT 926
      ::::::::::: ||| ||||| ||||| :::::::::::
796 aSerGIleProGIleProGIleProGIleProGIleProGIleProGIleProGIle 813
      ::::::::::: ||| ||||| ||||| :::::::::::
927 CAGCAATATCTTTATGCGACCATCCCAAGGATTCGAGCTGCTG 976
      ::::::::::: ||| ||||| ||||| :::::::::::
813 eGIleProGIleProGIleProGIleProGIleProGIleProGIleProGIle 829
      ::::::::::: ||| ||||| ||||| :::::::::::
977 GGGGAAATACGCTTGGCGGCAATCA.....CGGACATCT 1014
      ::::::::::: ||| ||||| ||||| :::::::::::
830 GlnIleProGIleProGIleProGIleProGIleProGIleProGIleProGIle 846
      ::::::::::: ||| ||||| ||||| :::::::::::
1015 GTCAAGCGCTGCGAGATGCGCGCATGCGATTCGCGGAAAGAAATCCG 1064
      ::::::::::: ||| ||||| ||||| :::::::::::
846 rGIleProGIleProGIleProGIleProGIleProGIleProGIleProGIle 863
      ::::::::::: ||| ||||| ||||| :::::::::::
1065 GTCGCGACGATTTGCGGATGCGGATACGCGCAATACCGCTCCCT 1114
      ::::::::::: ||| ||||| ||||| :::::::::::
863 rGIleProGIleProGIleProGIleProGIleProGIleProGIleProGIle 879
      ::::::::::: ||| ||||| ||||| :::::::::::
1115 ACCATTCGCAATATCGCT.....CAACTTGGACGACGCT 1152
      ::::::::::: ||| ||||| ||||| :::::::::::
880 ThrIleProGIleProGIleProGIleProGIleProGIleProGIleProGIle 896
      ::::::::::: ||| ||||| ||||| :::::::::::
1153 TACGGCAAGAAACATGCTCTCAACGCTGCGCGGCAACG... 1199
      ::::::::::: ||| ||||| ||||| :::::::::::
896 oHISerProGIleProGIleProGIleProGIleProGIleProGIleProGIle 912
      ::::::::::: ||| ||||| ||||| :::::::::::
1200 .....CAAAATGTCAACTGCGCA..... 1220
      ::::::::::: ||| ||||| ||||| :::::::::::
913 AlaGIleProGIleProGIleProGIleProGIleProGIleProGIleProGIle 929
      ::::::::::: ||| ||||| ||||| :::::::::::
1221 .....CCAGCGCACC 1232
      ::::::::::: ||| ||||| ||||| :::::::::::
929 uAlaSerGIleProGIleProGIleProGIleProGIleProGIleProGIleProGIle 945
      ::::::::::: ||| ||||| ||||| :::::::::::

```

```

seq_name: sp_human:Q9H487
seq_documentation_block:
ID Q9H487 PRELIMINARY; PRT: 1199 AA.
AC Q9H487;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUC4 PROTEIN SPLICED VARIANT SV15.
GN MUC4.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREATIC TUMOUR;
RA Choudhury A., Montano N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBD databases.
DR EMBL: A1400852; CAC14137.1;
SQ SEQUENCE 1199 AA: 123130 MW: 136C8F6977E71B CRC64;

```

```

alignment_scores:
  Quality: 155.50      Length: 517
  Ratio: 0.713        Gaps: 21
  Percent Similarity: 42.166      Percent Identity: 22.631
alignment_block:
US-09-303-518d-463 x Q9H487
Align seg 1/1 to: Q9H487 from: 1 to: 1199

```

```

9 TTCCCGCAAAATATCCCTATTCTGTCCATAC..... 41
493 trpProSerSerPheSerSerIysGlyHisThrThrTrpSerGlnThrG1 509
42 ...GGCAGTGTGCTGCCGATGCATGCACACGCCCTCAGATTGGCAAC 87
509 uLeuProSerThrSerThrGlyAlaIleThrArgLeuValThrGlyAsn 526
88 GATCCCTTATCCGGCAGGTCTGTGACCGTC..... 118
556 roSerThrArgAlaIleGlyThrIleProArgValProSerIysValSer 542
119 AGCATTTTCGACCCGACGCGGAATATCCACCTATTCGGACGAGCGGGAG 168
543 AlaIleGlyIleProGlyIleProThrThrIysSerSerHisSerThr 559
169 CTTCGCNAGCCGCAACG.....GCCAATCCGGAATGG..... 199
559 rLeuProIysThrThrGlyAlaGlyAlaGlnThrGlnThrGlnIle 576
200 .....GAAACAATAC 208
576 hrGlyThrThrGlyAlaLeuLeuSerSerProSerIysSerValIle 592
209 AAAGCCATCAGTTGGCCACCTGATGATTCACAGCGCGCTTG...AA 255
593 GlnMetIleIysThrAlaThrSerProSerSerSerProMetLeuAsp 609
256 GGAATATCCGCTACATTCGCTTTCCGATCAGCGGACAAATTCGA 305
609 gHisThrSerGlnIleThrThrAlaProSerThrAsnHisSerThrI 626
306 TTCGCCCTTGACAAACATGCCT...CACATCCGATTCG..... 343
626 lHisSerThrSerThrSerProGlnIleSerProAlaValSerGlnArg 642
344 .....ACGAAGCC 351
643 GlyHisThrArgAlaProGlnThrThrGlnIleSerGlnThrThrArg 659
352 GGTATGCCCGTTCGATTCAGCTTACCAGCATTCATGGACGCGATA 401
659 rValSerProMetThrAsp.....ThrIysT 668
402 CGAACACCATCCCGCGCGCGCTATGACG.....GGCCACAGGCGCG 445
668 hrValThrThrProGlySerSerPheThrAlaSerGlyHisSerPro 684
446 GCTATCCCGCTCCCAAGCGCGAGGATATATACAGTACAGATMAA 495
685 GluIleValProGlnAspAla..... 691
496 GCGGTTCGCCAAATATCCGCTCAACCTGACCGACACCGCAGC 545
692 .....ProThrIleSerAlaIleThrThrPheAlaProAlaProThrG 706
546 ACAAGCGCTTGGCAGCGCTTCCCAATGCGCGCTATGATGACGCAAG 595
706 LysAsnGlyHisThrThrGlnAlaProThrThrAlaLeu..... 718
596 GAGTAGCGCGAGTTCMAAGCGCGACCGCATACGCCCGAGCTGAGC 645
719 .....GlnAlaIleProSerSerHis 725
646 AGATGGCGCAATGCCCGAAGCCTTCAACGCGACGACGATATGCTCA 695
725 sAspAlaThrLeuGlyProSerGlyIleThrSerLeu.....SerL 739
696 AAACATCATTCGCGCGCGAGGAAATTCGCGCGACGCGATGCCGCTC 745
739 ySerThrGly..... 741

```

```

746 AGGGTATTAAGCGAAGCCTCAACATTTGCTGTCTATGCACGCGTGGCTG 795
742 .....AlaLeuThrIleAlaAlaAsnSerValValSerThrPr 753
796 CTTTCACCGCAAAACAAGATGGCGGCATCCACGATTTGGCAGATGGC 845
753 ogIleGlyProGlnGlyIleThrThrSerAla..... 763
846 GCAACTCAAAAGACTATCCCGCAGCAGCAGCATCCGCGAT..... 883
764 ..SerAlaSerThrSerProAspThrAlaAlaIleThrHisThrHis 779
884 .....GGCAGTCCAAACCCCAATGCCGA 909
780 GlnAlaGlySerThrGlnAlaIleSerGlyIleThrGlnThrSerGlnProAl 796
910 CAAGCA.....TAGAAGCGT 926
796 aserSerGlySerArgThrThrSerAlaGlyThrAlaThrProSerSerS 813
927 CAGCAATATCTTTATGCGACCGATCCCATCAAGAGATTCGAGCTGCC 976
813 erGlyAlaSerGlyThrThrProSerGlySerGlyIleThrSerThr 829
977 GGGGAAATACGGCTTGGCGGCATCA.....CGGCACATCCT 1014
830 GlyIleThrThrArgPheSerSerAsnProSerArgAspSerHisThrTh 846
1015 GTCACACGCTCCGACATGGCGCGCATCCGATCCGGAAGGAAATCCGC 1064
846 rGlnSerThrThrGlnLeuLeuSerAlaSerAlaSerHisGlyAlaIleP 863
1065 CGTCACGCGCAATTTCCGATGCGCGCATACGCCAAATACCCGCTCCCT 1114
863 rValSerThrGlyMetAlaSerSerIleValProGlyThrPheHisPro 879
1115 ACCATTCGCCAAATATCCGCT.....CAACTGTGGACGAGCT 1152
880 ThrLeuSerGlnAlaSerThrThrAlaGlyArgProThrGlyGlnSerSerPr 896
1153 TACGCGCAAGAAACATCATCTCTCAACCGTCCGCGCTCAACGCG... 1199
896 oThrSerProSerAlaSerProGlnGluThrAlaAlaIleSerArgMet 912
1200 .....CAAAATGTCAACTGGCAGA..... 1220
913 AlaGlnThrGlnArgThrGlyThrSerArgIleSerAspThrIleSerLe 929
1221 .....CCACGCGCCAGCC 1232
929 uAlaSerGlnAlaThrAspThrPheSerThrValProProThrProPro 945

seq_name: sp_human:09H486
seq_documentation_block:
ID 09H486 PRELIMINARY; PRT; 1201 AA.
AC 09H486:
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUC4 PROTEIN SPLICED VARIANT SV16.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREATIC TUMOUR.
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ400853; CAC14138.1; -

```


GN MUC4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PANCREATIC TUMOUR;
 RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
 RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma."
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ400854; CAC14139.1;
 SO SEQUENCE 1214 AA; 125245 MW; B89D6A84AD4A056 CRC64;

alignment_scores:
 Quality: 155.50 Length: 517
 Ratio: 0.713 Gaps: 21
 Percent Similarity: 42.166 Percent Identity: 22.631

alignment_block:

US-09-303-518D-463 x Q9H485 ..

Align seg 1/1 to: Q9H485 from: 1 to: 1214

```

9  TTCCCGCAAAATATCCCTTATCTGTCCACT..... 41
   ::::::::::::::: ::::
493  TrpProSerSerPheSerSerIysGlyHisThrThrTrpSerGlnThrG1 509
   ::::::::::::::: ::::
42  ....GGCAGTGTGCTGCCGATGCATGCACAGCGCTCAGATTGGCAAC 87
   ::::::::::::::: ::::
509  uIeuProSerThrSerThrGlyAlaAlaThrArgIeuValThrGlyLysn 526
   ::::::::::::::: ::::
88  GATCCCTTATCCGCGAGTTCCTGACCGTC..... 118
   ::::::::::::::: ::::
526  roserThrArgAlaAlaGlyThrIleProArgValProSerLysValSer 542
   ::::::::::::::: ::::
119  AGCATTTTGAACCCGACGGAAATACACCATTCATCGCAGCAGGAGGAG 168
   ::::::::::::::: ::::
543  AlaIleGlyIuProGlyIuProThrThrTyrSerSerHisSerThr 559
   ::::::::::::::: ::::
169  CTGCGCAAGCGCAGC.....GCCATTCGCGATTG..... 199
   ::::::::::::::: ::::
559  rIeuProLysThrThrGlyAlaGlyAlaGlnThrGlnThrGlnLut 576
   ::::::::::::::: ::::
200  .....GAAACATAC 208
   ::::::::::::::: ::::
576  hrGlyThrThrGlyAlaAlaIeuLeuSerSerProSerTyrSerValIle 592
   ::::::::::::::: ::::
209  AAAGCCATCAAGTTGGCCACCTGATGATCAACAGCGCGCGTTG...AA 255
   ::::::::::::::: ::::
593  GlnIleIleLysThrAlaThrSerProSerSerSerProMetIeuSpar 609
   ::::::::::::::: ::::
256  GGAATATCGGCTCATTTCCGCTTTCCGATCAGCGGACAAATTCGA 305
   ::::::::::::::: ::::
609  gHisThrSerGlnIleThrThrAlaProSerThrAsnHisSerThrI 626
   ::::::::::::::: ::::
306  TTCGCCCTTCGACAAACCATGCT...CACATTCGATTGTG..... 343
   ::::::::::::::: ::::
626  lHisSerThrSerThrSerProGlnIuSerProAlaValSerGlnAlu 642
   ::::::::::::::: ::::
344  .....ACGAAGCC 351
   ::::::::::::::: ::::
643  GlyHisThrArgAlaProGlnThrThrGlnIuSerGlnThrThrArgse 659
   ::::::::::::::: ::::
352  GGTAGTCCCGTTCGAGGATTCAGCCTTTACCGCATCATTGGACGGATA 401
   ::::::::::::::: ::::
659  rValSerProMetThrAsp.....ThrLysr 668
   ::::::::::::::: ::::
402  CGAACACCATCCGCGCGACGGCTATGACG.....GGCCACAGCGCGGG 445
   ::::::::::::::: ::::
668  hrValThrThrProGlySerSerPheThrAlaSerGlyHisSerProser 684

```

```

446  GCTATCCCGCTCCCAAGCGCGAGGGATATATACGCTATGACATPAAAA 495
   ::::::::::::::: ::::
685  GlnIleValProGlnAspAla..... 691
   ::::::::::::::: ::::
496  GGCCTGCCCAAAATATCCGCTTCACCTGACGCGACAGCAGCAGCCGG 545
   ::::::::::::::: ::::
692  .....ProThrIleSerAlaAlaThrThrPheAlaProAlaProThrG 706
   ::::::::::::::: ::::
546  ACAACGGCTTCGCGACCGCTTCACATGCGCGCTATGCTGACGCAAG 595
   ::::::::::::::: ::::
706  LysnGlyHisThrThrGlnAlaProThrThrAlaIeu..... 718
   ::::::::::::::: ::::
596  GAGTAGCGCAGCGATTCAAACGCGCCACCGCATACAGCCCGCGCTGAC 645
   ::::::::::::::: ::::
719  .....GlnAlaAlaProSerSerH 725
   ::::::::::::::: ::::
646  AGATCGCGCAATGCGCGCGAAGCCTTCACGCGCTGACATTCGTCAA 695
   ::::::::::::::: ::::
725  sAspAlaThrLeuGlyProSerGlyGlyThrSerLeu.....SerL 739
   ::::::::::::::: ::::
696  AAACATCATTCGCGCGCGACGAGAAATTCGTGCGCGACGCGATGCCGTG 745
   ::::::::::::::: ::::
739  ySThrGly..... 741
   ::::::::::::::: ::::
746  AGGTATATAGCGAAGCTCAACATTCGTCTCATGACGCGCTGGGTCTG 795
   ::::::::::::::: ::::
742  .....AlaLeuThrIleAlaAsnSerValValSerThrPr 753
   ::::::::::::::: ::::
796  CTTTCACGCGAAACAAAGATGGCGCGCATCAACGATTGGAGATATGCG 845
   ::::::::::::::: ::::
753  oGlyGlyProGlnGlyGlnThrPheSerAla..... 763
   ::::::::::::::: ::::
846  GCACATCAAGACTATGCGCGACGACCATTCGCGATT..... 883
   ::::::::::::::: ::::
764  ..SerAlaSerThrSerProAspThrAlaAlaIleThrHis 779
   ::::::::::::::: ::::
884  .....GGCGAGTCCCAAAACCCCATGCGCGA 909
   ::::::::::::::: ::::
780  GlnAlaGlnSerThrGlnAlaSerGlyGlnThrGlnThrSerGlnProAl 796
   ::::::::::::::: ::::
910  CAAGGCA.....TAGAAGCGCT 926
   ::::::::::::::: ::::
796  aserSerGlySerArgThrThrSerAlaGlyThrAlaThrProSerSer 813
   ::::::::::::::: ::::
927  CAGCAATATCTTATGCGACGCATCCCATCAAAAGGATTCGACGCTGCC 976
   ::::::::::::::: ::::
813  erGlyAlaSerGlyThrThrProSerGlySerGlyIleSerThrSer 829
   ::::::::::::::: ::::
977  GGGAAATATACGCTTGGCGGCGCATCA.....CGGCACATCCT 1014
   ::::::::::::::: ::::
830  GlyGlnThrThrArgPheSerSerSerProSerArgAspSerHisThr 846
   ::::::::::::::: ::::
1015  GTCAAAGCGTCCAGATGGCGCGCATTCGATTCGCGAAAGGAAATCCGC 1064
   ::::::::::::::: ::::
846  rGlnSerThrThrGlnIleuLeuSerAlaSerAlaSerHisGlyAlaIle 863
   ::::::::::::::: ::::
1065  CGTCAGCGCAATTTTGGCGATGCGCGCATACGCCAAATACCGCTCCCTT 1114
   ::::::::::::::: ::::
863  rValSerThrGlyMetAlaSerSerIleValProGlyThrPheHisPro 879
   ::::::::::::::: ::::
1115  ACCATTCGCGAAATATCCGTT.....CAACCTTGAGCAGCGCT 1152
   ::::::::::::::: ::::
880  ThrIeuSerGlnAlaSerThrAlaGlyArgProThrGlnGlnSerSerPr 896
   ::::::::::::::: ::::
1153  TACGGCAAAAGAAACATACCTCTCAACCGTGGCGCGTCAAAAGG... 1199
   ::::::::::::::: ::::
896  oThrSerProSerAlaSerProGlnGluThrAlaAlaIleSerArgmet 912
   ::::::::::::::: ::::
1200  .....CAAAATGTCAAACTGCGACA..... 1220
   ::::::::::::::: ::::
913  AlaGlnThrGlnArgThrGlyThrSerArgGlySerAspThrIleSerLe 929
   ::::::::::::::: ::::
1221  .....CCAAGCGCACCC 1232

```

929 uAlaSerGlnAlaThrAspThrPheSerThrValProProThrProPro 945

seq_name: sp_human:Q9H484

seq_documentation_block:

ID Q9H484 PRELIMINARY; PRT; 1215 AA.

AC Q9H484:

DT 01-MAR-2001 (TRENBLER, 16, Created)

DT 01-MAR-2001 (TRENBLER, 16, Last sequence update)

DT 01-JUN-2001 (TRENBLER, 17, Last annotation update)

DE MUC4 PROTEIN SPLICED VARIANT SV18.

GN MUC4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-PANCREATIC TUMOUR;

RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;

RT Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.

RL EMBL: AJ400855; CAC14140.1;

DR InterPro: IPR003886; NIDogen_ext.

DR SMART: SM00539; NIDO; 1.

SEQUENCE 1215 AA; 125435 MW; 2B950CD46945D60D CRC64;

alignment_scores:

Quality: 155.50 Length: 517

Ratio: 0.713 Gaps: 21

Percent Similarity: 42.166 Percent Identity: 22.631

alignment_block:

US-09-303-518D-463 x Q9H484 ..

Align seg 1/1 to: Q9H484 from: 1 to: 1215

```
9 TTCCCGCAAAATATCCCTTATCTGTCATATC..... 41
:::|||||:::
493 TTPProSerSerPheSerSerLysGlnHisThrThrTrpSerGlnThrG 509
42 ....GGCAGTGTGCTCCGATCGATGACACAGCCCTCAGATTGGCAAC 87
:::|||||:::
509 uLeuProSerThrSerThrGlnAlaAlaThrArgLeuValThrGlyAsn 526
88 GATCCCTTATCCGGCAGTCTCGACCGTC..... 118
:::|||||:::
526 roSerThrArgAlaAlaGlyThrLeuProArgValProSerLysValSer 542
119 AGCATTTGCAACCGAGCGGAAATACCACTATTCGGACAGCGGGGAG 168
:::|||||:::
543 AlalleGlyGluProGlyGluProThrThrLysSerSerHisSerThr 559
169 CTTCGCCAAGCGCAAG.....GCCATATCGATTG..... 199
:::|||||:::
559 rLeuProLysThrThrGlyAlaGlyAlaGlnThrGlnTrpThrGlnGlu 576
200 .....GAACATAC 208
576 hrGlyThrThrGlyAlaAlaLeuLeuSerSerProSerThrValIle 592
209 AAAGCCATGATGGCCACCTGATGATCAACAGCGGCGCTTG...AA 255
:::|||||:::
593 GlnMetLeuThrAlaThrSerProSerSerProMetLeuAsp 609
256 GGAATATCGGCTACATGTCGCTTTCGATCAGCGGCAAAATTCGA 305
:::|||||:::
609 gHisThrSerGlnGlnThrThrAlaProSerThrAsnHisSerThr 626
306 TTCCGCTTGCAGAACCATGCT...CACATTCGATTCTG..... 343
:::|||||:::
::: ||| :::
```

```
626 LeHisSerThrSerThrSerProGlnGluSerProAlaValSerGlnArg 642
344 .....ACGAGCC 351
643 GlyHisThrArgAlaProGlnThrThrGlnGluSerGlnThrThrArg 659
352 GGTAGTCCCGTGAAGGATTCAGCCTTACCAGCATCATGGAGGATA 401
659 rValSerProMetThrAsp.....ThrLysT 668
402 CGAACACCATCCCGCCGAGCTATGAG.....GGCCACAGCGGGCG 445
668 hrValThrThrProGlySerSerPheThrAlaSerGlyHisSerPro 664
446 GCTATCCCGCTCCCAAGCGCGAGGATATACAGCTACGACATAAA 495
685 GluLeuValProGlnAspAla..... 691
496 GGCCTTGGCCAAATATCCGCTTCACCTGACGACACCGCAGCAGCG 545
692 .....ProThrLysSerAlaAlaThrThrPheAlaProAlaProThr 706
546 ACAAGGCTTCCGAGCCGTTTCCAAATGCCGCGCTATGCTGACGCA 595
706 LysGlnGlyHisThrThrGlnAlaProThrThrAlaLeu..... 718
596 GAGTAGCGCAGGATTCAAACGCCGCCACCGATACAGCCCGAGCTG 645
719 .....GlnAlaAlaProSerSerH 725
646 AGATCGGCAATGCGCGAGCGCTTCACGCGCAGCTACAGATATGCA 695
725 SASpAlaThrLeuGlyProSerGlyGlyThrSerLeu.....SerL 739
696 AAACATCATCGCGCGGCGAGCAAAATGTGCGGCGAGCGAGCTGCG 745
739 ysthGly..... 741
746 AGGATATAGCGAAGGCTCAACATGTGTCATGCAAGCGCTGGGTG 795
742 .....AlaLeuThrLeuAlaAsnSerValValSerThrPr 753
796 CTTCACCGCAAAACAGATGCGCGCATCAACGATTTGGCAGATATG 845
753 oglyGlyProGluGlyGlnTrpThrSerAla..... 763
846 GCAACTCAAAAGCTATGCGCAGCAGCATCCGCGAT..... 883
764 ..SerAlaSerThrSerProAspThrAlaAlaAlaMetThrHis 779
884 .....GGGAGTCCAAACCCCAATGCCGA 909
780 GlnAlaGluSerThrGlnAlaSerGlyGlnThrGlnThrSerGluPro 796
910 CAAGCA.....TAGAGCGT 926
796 aserSerGlySerArgThrThrSerAlaGlyThrAlaThrProSer 813
927 CAGCAATATCTTATGAGCAGCATCCCATCAAGGATTTGGAGTGGC 976
813 erGlyAlaSerGlyThrThrProSerGlySerGlyLysThrSer 829
977 GGGAAATACGCTTGGCGCGCATCA.....CGGCACATCCT 1014
830 GlyGluThrThrArgPheSerSerAsnProSerArgAspSerHisThr 846
1015 GTCAAGCGCTCCAGATGAGCGCATTCGATTCGCAAGGAAATCCGC 1064
846 rGlnSerThrThrGlyLeuLeuSerAlaSerAlaSerHisGlyAla 863
1065 CGTCAGCGCAATTTGGCGATGCGGATACGCCCAAAATGCCGCTT 1114
863 roValSerThrGlyMetAlaSerLeuValProGlyThrPheHisPro 879
```



```

||||| ||||| ||||| ||||| |||||
813 etclvalaserghlythrphroserglsergllyleaserhser 829
977 GGGGAAATATCGGCTTGGGGGATCA.....CGGCACATCCT 1014
||||| ||||| ||||| ||||| |||||
830 gylglthrthrargpheserlaserproserargaspserrhthrth 846
1015 GTCAGCGGTCCGACATGGCGCCGATCGCATCGCGAAGGAATCCGC 1064
||||| ||||| ||||| ||||| |||||
846 rglaserthrthrghluleuseralaserlaserhsglyalaller 863
1065 CGTCAGCGCAATTTGGCGATGGCGCATCGCAAAATACCGCTCCTT 1114
||||| ||||| ||||| ||||| |||||
863 rovalserthrghlymetlaserlaserlilevalproglthrphenspro 879
1115 ACCATTCGCAAAATATCGCTT.....CAAACTTGAGACAGCGT 1152
||||| ||||| ||||| ||||| |||||
880 thrleuserghlualaserthrarglthrprothrghlylasersepr 896
1153 TAGCGCAAGAAACATCATCTCCATACCGTGGCGCGTCAACGCG... 1199
||||| ||||| ||||| ||||| |||||
896 othrserproseralaserproglthrthrAlalalaserargmet 912
1200 .....CAAAATGTCAAACTGGGAGA..... 1220
913 Alaglnthrghlthrghlythrserarglserasprthrleaserle 929
1221 .....CGACGCCAGCC 1232
929 uAlaserghlAlathrasprthrpheserthrvalprothrpropro 945

```

seq_name: sp_human:Q9H482

```

seq_documentation_block:
ID Q9H482; PRELIMINARY; PRT; 1827 AA.
AC Q9H482;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MUC4 PROTEIN SPLICED VARIANT SV20.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Carnivora; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TRISSE-PANCREATIC TUMOUR;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ008857; CAC14142.1; -
DR InterPro: IPR003886; Nidogen_ext.
DR InterPro: IPR003662; sub_transporter.
DR InterPro: IPR001846; Vwd.
DR Pfam: PF00094; vwd; 2.
DR SMART: SM00539; NIDO; 1.
DR SMART: SM00216; VWD; 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
SQ SEQUENCE 1827 AA; 194100 MW; 63CE7895C0F8741 CRC64;

```

alignment_scores:

```

Quality: 155.50 Length: 517
Ratio: 0.713 Gaps: 21
Percent Similarity: 42.166 Percent Identity: 22.631

```

alignment_block:

us-09-303-518d-463 x Q9H482 ..

Align seg 1/1 to: Q9H482 from: 1 to: 1827

9 TTCCCGCAAAATATCGCTTATTCGTGCATACT..... 41

```

493 TprroserpheserlserlsglyhthrthrTprserlthrgl 509
42 ....GGCAGTGTGCTGCTGCATGCACAGCGCTCAGATTGGCAAC 87
509 uleuproserthrserthrghlAlalathrghlvalthrghlyasn 526
88 GATCCCTTTATCCGCGAGTTCTGACCTC..... 118
526 roserthrargAlalaglythrleproargvalproserlvalser 542
119 AGCATTTGAAACCCGAGCGGAATACCCATTTGCGACAGGGGAG 168
543 Alalleglthrproglthrprothrthrlyserlserhserthrth 559
169 CTGCGCAGGCGCAACG.....GCCATATCGGATTGG..... 199
559 rleuprolythrthrghlyalaglyalaglnthrghlthrghlthr 576
200 .....GAAACATVAC 208
576 hrghlythrghlyalalaleuleuserlaserproserlthrvalle 592
209 AAAGCATCATGTCGGCCACCTGATGATTCACAGGGCGCTTG...AA 255
593 GlnmetlthrAlathrthrserproserlaserproleuaspar 609
256 GGAATATCGGCTACATTTGCGCTTTGATCATCGGCGCAAAATTCGA 305
609 ghlsrthrserlthrghlthrthrAlalaproserthrAsnhlsrthr 626
306 TTGCGCCCTTGCAACCATGCT...CACATTCGATTCTG..... 343
626 lehlserthrserthrserproglthrproAlalalaserghlthr 642
344 .....ACGAGCC 351
643 ghlythrargAlalproglthrthrghlthrghlthrghlthrthr 659
352 GGTAGTCCGCTGAGCGATTCAGCTTACCGCATTCATTCGAGGATA 401
659 rvalserprolethrserp.....Thrlyst 668
402 CGAACCATCCCGCGCGAGCTATGACG.....GGCCACAGGGCGCG 445
668 hrvalthrthrproglthrserlserphelthrAlalaserghlthr 684
446 GGTATCCCGCTCCAAAGCGCGAGGATATATACGCTATACATATAA 495
685 ghulthrvalproglthrAlal..... 691
496 GCGCTTGCCCAAAATATCCGCTGCAACCTGACGACACCGACCGG 545
692 .....ProthrleaserlAlalathrthrRhealproAlalprothr 706
546 ACACAGCGCTTCCGACCGCTTTCACAAATCCGGCGCTATGCTACG 595
706 lylasnghlyhthrthrghlAlalprothrthrAlalaleu..... 718
596 GAGTAGCGGAGGATTCAAAGCGCGCACCGATACAGCGCCGAGCTG 645
719 .....GlnAlalalaproserlserh 725
646 AGATCGGCAATCGCGCGAAGCCTTTCACAGCGCATGATTCGTA 695
725 saspralthrleughlyproserghlythrserleu.....Serl 739
696 AAACATCATCGCGCGCGACGAGAAATGTGCGGCGACGCGATGCG 745
739 ysfhrghly..... 741
746 AGGTATATAGCGAGCTCAACATTCGTGCATGACGCGCTGGCTG 795

```


[illegible]

```

1065 CGCAGCGCAACAATTTCGGCCATGCGGCATACGCGCAAAATACCCGTCCTT 1114
      ||::|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
863 rovalSerThrGlyMetAlaSerSerIleValProGlyThrPheHisPro 879
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1115 AACATTCGCGAAATATCGCTT.....CAAACTGGAGCAGCT 1152
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
880 ThrIleSerGluAlaSerThrAlaGlyArgProThrGlyGlnSerSerP 896
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1153 TAGCGCAAGAAAGAAATCATCCCTCTCTCAACCGCGCCGCGTCAAAAG... 1199
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
896 thrSerProSerAlaSerProGlnGlu.ThrAlaAlaIleSerArgMet 912
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1200 .....CAAAAATGTCAACTGGCAGC..... 1220
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
913 AlaGlnThrGlnArgThrGlyThrSerArgGlySerArgThrIleSerIe 929
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1221 .....CCAGCGCACCC 1232
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
929 uAlaSerGlnAlaThrAspThrPheSerThrValProThrProPro 945
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
seq_name: sp_human:Q9NY09

seq_documentation_block:
ID Q9NY09 PRELIMINARY; PRT; 2169 AA.
AC Q9NY09;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE MUCIN 4.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
[1]
RX MEDLINE=20381033; PubMed=10920259;
RA Choudhury A., Moniaux N., Wimpenny J.P., Hollingsworth M.A.,
RA Audert J.P., Batra S.K.;
RA "Human MUC4 mucin cDNA and its variants in pancreatic carcinoma."
RL J. Biochem. 128:233-243(2000).
EMBL: AJ276359; CAB81773.1; -.
InterPro: IPR000561; EGF-like.
InterPro: IPR003866; Nidogen_ext.
InterPro: IPR003662; sub-transposr.
DR Pfam: PF00094; vwd: 2.
DR SMART: SM00181; EGF, 1.
DR SMART: SM00539; NIDO; 1.
DR SMART: SM00216; VWD; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
DE SEQUENCE 2169 AA; 231470 MW; 3CBAB16AF73E469 CRC64;

```

```

alignment_scores:
    quality: 155.50      Length: 517
    Ratio: 0.713        Gaps: 21
    Percent Similarity: 42.166      Percent Identity: 22.631

alignment_block:
US-09-303-518D-463 x Q9NY09 ..

Align seg 1/1 to: Q9NY09 from: 1 to: 2169

9 TTCGCGCAAAATATCCCTTATCTGTCCAACT..... 41
:::||||:||||:
493 TrpProSerSerPheSerSerIyGlnIshThrTrpSerGlnIhTrI 509
:::||||:||||:

42 ... GGCAGTGTGCTGCCAGATGCATGCACACGGCTCAGATTGGCAAC 87
||||:||||:||||:||||:
509 uIeudProSerTherTrIyAlaAlaIhTrArgLeuValThfGlyAsnP 526

```

88	GATCCCTTAAATCCGGCAGTTCTGCAGCGTC	118
526	rosethrhtrgalaalaaaglynhtrlperogayalproserlyvalser	542
119	AGCATTTCCGAAACCCGACGAGAAATACCAACCATTTCCGACGAGGGGAG	168
543	AlailegylgubProglyolubProthrhthtyserserthserthrh	559
169	CTTGGCCNACGGCAAGC.....GCCATATCGGATTCG.....	199
559	rlenProlystrhtrhglaaglyalaglnlthrlntprthglnlgt	576
200GAAACATAC	208
576	hrclyhtrhtrhgylualaleuleuserserproserlyrserValle	592
209	AAAGCCATAGTTGGGCCACCTGATGATTCAACAGCGCCGCTTG..AA	255
593	GlmetlelystrhtrhtrhserProserSerSerPrometleuaspAr	609
256	GGAAATATAGGGTACATATGTCCGGCTTTCCATGACGGGACAATTCGA	305
609	ghlstrhserglnlntlthtrhtrhAlproserThAsnhlserThrt	626
306	TTGCGCCTTCGACAAACCATTCCT..CACATTCGGATTCG.....	343
626	lenlserhtrserhtrhserProglnglnuserProalValserGlnatg	642
344ACGAGCC	351
643	glnhtrhtrgalaalrproglntlrhtrhglnluserglntrhtrhtrngse	659
352	GGTAGTCCCGGTGACGAGATTCAGCCTTTACCGCATTCGACGAGGAGTA	401
659	rlalserPrometleuasp.....Thlysn	668
402	CGAACACCATCCCGCCGACGGCTAGACG.....GGCCACAGGGCGGCG	445
668	htrValhtrhtrProglySerSerPhetrlhAlaserGlnhtrhtrhserProser	664
446	GTATCCCGCTCCCAAGGCGCGAGGATATATACAGTACGACATATAA	495
685	ghlulealrProglaspAla.....	691
496	GGCGTTCCCAAAATATTCGCGCTCAACTACGACCGACACGACACCGG	545
692ProthrllesterAlaAlaThtrhPhelaAlrAlaProthrg	706
546	ACAAAGCGCTCGGACGCGTTCCCAAAATCCGCGCGCTATGCTGACGACG	595
706	lyasnlglnhtrhtrhglnalrAlaProthrhtrhAlaleu.....	718
596	GAGTAGCGCAGGATTCAAACGCGCACCCGATACAGCCCGACCTTGAC	645
719GlnAlaAlaProserSerh	725
646	AGATCGGGCAATCCCGCCGACGCTTCAACGCGACATGAGATATGTCGA	695
725	sAspAlaAlrleuGlyProserGlylylthtrserleu.....serL	739
696	AAATCATCATCGGCGCGGACGAGAAATGTGCGGCGACGATCCCGTGC	745
739	ytshrgly.....	741
746	AGGGTATTAACCGAAGCTCAAAACATTTGCTGCATGACGCGTTGGGTCG	795
742AlaleuThrlleuAlaAsnSerValAlaserThtrPr	753
796	CTTTCCACCGAAACAAAGATGGCGGCGCATCAACGATGGCAGATATAGGC	845
753	oglyGlyProglnglyGlntrhtrhserAla.....	763

```

846 GCAATCTCAAGACATGATGCGGACGACCATCGGGAT..... 883
      ..:::..|||:::..:::..|||
764 ..SerAlaSerThrSerProAspThrAlaAlaAlaMetThrHis779
884 .....GGGCAAGTCCAAAACCCCAATGCCGCA 909
      |||:::..|||
780 GlnAlaGluSerThrGluAlaSerGlyGlnThrGlnThrSerGluProAl 796
      |||:::..|||
910 CAAGCA.....TAGAGCGGT 926
      :::..:::..
796 AserSerIlySerAlaThrThrSerAlaGlyThrAlaThrProSer's 813
      :::..:::..
927 CAGCATATCTTTATGCGACCATCCCATCAAGAGATGGAGCTGTCC 976
      |||:::..|||
813 erGlyAlaSerGlyThrThrProSerGlySerGlyGlyIleSerThrSer 829
      |||:::..|||
977 GGGGAAATACGGCTTGGGGGCGATCA.....CGGCACATCCT 1014
      |||:::..|||
830 GlyGluThrThrArgPheSerSerSerProSerArgAspSerHisThrTh 846
      |||:::..|||
1015 GTCAAGCGCGTGCAGATGGGCGCGATCCATTCGCGCAAGAGAAATCCGC 1064
      :::..|||
846 rGlnSerThrThrGlyLeuLeuSerAlaSerAlaSerHisGlyAlaIleP 863
      :::..|||
1065 CGTCAGCAGCAATTTTGGCGATGCGGCGATACGCCAATACCCGTCCTT 1114
      |||:::..|||
863 rovalSerThrGlyMetAlaSerSerIleValProGlyThrPheHisPro 879
      :::..|||
1115 ACCAATCCCGAAATATCGGT.....CAACTGTGACGACGCT 1152
      |||:::..|||
880 ThrLeuSerGlnAlaSerThrAlaGlyArgProThrGlyGlnSerSerPr 896
      |||:::..|||
1153 TAGGCGCAAAAGAAACATCAGCTCTCTCAACGCGCGCGCTCAACG... 1199
      |||:::..|||
896 oThrSerProSerAlaSerProGlnGlu..rThAlaAlaIleSerArgMet 912
      |||:::..|||
1200 .....CAAAATGTCAAACTGGCGCA..... 1220
      |||:::..|||
913 AlaGlnThrGlnArgThrGlyThrSerArgGlySerAspThrIleSerIle 929
      |||:::..|||
1221 .....CCAAAGCGCACCC 1232
      |||:::..|||
929 uAlaSerGlnAlaThrAspThrPheSerThrValProProThrProPro 945
      |||:::..|||

seq_name: sp_human:Q14881

seq_documentation_block:
ID Q14881 PRELIMINARY; PRT; 622 AA.
AC Q14881.
DT 01-NOV-1996 (TREMBlrel_01, Created)
DT 01-NOV-1996 (TREMBlrel_01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel_19, Last annotation update)
DE MUCIN (FRAGMENT).
GN MUC5B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TRACHEOBRONCHIAL MUCOSA;
RA Dessevay J.L., Guyonnet-Duperrat V., Porchet N., Aubert J.P., Laine A.;
RT "Human mucin gene MUC5B: the 10.7 kb large central exon encodes
RT various alternate subdomains resulting in a super repeat.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
EMBL; X74955; CAA52910.1; -.
FT NON_TER 1
FT NON_TER 622
SQ SEQUENCE 622 AA; 61786 MW; 4FC85A52F50D57E5 CRC64;

alignment_scores:
Quality: 154.00 Length: 442

```

Ratio: 0.740 Gaps: 20
 Percent Similarity: 47.059 Percent Identity: 24.661
 Alignment block:
 US-09-303-518D-463 x Q14881 ..

Align seg 1/1 to: Q14881 from: 1 to: 622

```

68  ACGGCTCAGATTTGGCAACGATCCCTTTATCCGCGAGTTTCGACGCT 117
    |||||
187 ThrProGluThrThrHisThr.....SerThrVal 196
118 CAGCATTTCCAGACCCGAGGGAATACCATTTCCGAGCAGGCGGA 167
    |||||
196 IleuThrThrHisAlaThr.....ThrArgAlaThrSerSerHis 211
168 GCTTGGCCAGGCAAGGCCATATCGATGGGAAACATCAAAAGCCATC 217
    |||||
211 erThrProSerSerThrProGlyThrThrThrIleuThrGluLeuThr 227
218 AGTTGGCACCCTGATGATTCACAGCGCGCG..... 250
    |||||
228 ThrAlaAlaThrThrThrAlaGlyThrGlyProThrAlaThrProSer 244
251 .....TTGAAGAAATATCGCTACATTTGTCG 278
244 rThrProGlyThrThrThrThrIleuThrGluLeuThrThrAlaAla 261
279 CTTTTCGATCAGCGG.....ACAAATTCATTCGCTTCGACA 319
    |||||
261 hThrAlaSerThrGlySerThrAlaThrLeuSerSerThrProGlyThr 277
320 ACCATGCGCTGACATTCGATTTGACGAGCGGATGCGCTTGACGGA 369
    |||||
278 ThrTrp.....IleuThrGluProSerThrThrAlaThrVal 290
370 TTCAGCCTTTACGCGATCCATTGGAGCGATACGAACCATCCGCGCA 419
    |||||
290 lThrValProThrGlySerThrAlaThrAlaSerSerThr...GlnAla 366
420 CGGCTATGAGCGGCGACAGGCGCGGCTATCCGCTCCA..... 460
    |||||
306 hAlaGlyThrProHisValSerThrThrAlaThrThrProThrValThr 322
461 .....AAGCGCGAGGATATATACAGTACGACATAAAGCGTGGC 504
323 SerSerLysAlaThrProSerSerSerProGlyThrAlaThrAlaLeu 339
505 CAAATA...TCGCGCTCAACCTGACCGACA.....ACGCGAG 539
    |||||
339 oAlaLeuArgSerThrAlaThrThrProThrAlaThrSerPheThrAla 356
540 CACCGGACAGGCTTGCGGCGCT.....TCCACA 571
    |||||
356 leProSerSerSerLeuGlyThrThrThrArgLeuSerGlnThrThr 372
572 ATGCCGCGCTATGCTAGCGCAAGAGTAGGATTCAAACGCGCC 621
    |||||
373 ThrProThr.....AlaThrMetSerThrAlaThr 382
622 ACCGGATACAGCCCGAGCTGAGCAATGGGCAATCCGCGCAACCTT 671
    |||||
382 rProSerSerThrProGluThrVal.....HisThr 393
672 CAAGGCACTGCAGATATCGTCAAAACATCAACGCGCGGAGGAA 721
    |||||
393 erThrValLeuThrAlaThrAlaThrThrThrGlyAlaThrGlySerVal 409
722 TTTCGCGCAGGCGATGCCGTGACAGGATTAAGCGAAGCTCAACATT 771
    |||||
410 AlaThrProSerSerThrProGlyThrAlaAlaHisThrThrValPro 426
772 GCTGTATGAGCGGCTTGGGTCTGCTTCCACCGAAACAAATGCGCG 821
  
```

```

426 rThrThrThrThrGlyPheThrAlaThrProSerSerSerProGlyThrA 443
    |||||
822 CATCAACGATTTGGCAGATATGCGGCACTCAAGACTATGCG..... 865
    |||||
443 lAluThrPro...ProValTrpIleSerThrThrThrProThrThr 458
866 CAGCACCATCCGCGATTTGGCGATGCCAAACCCCATCCGCGACAGCG 915
    |||||
459 ThrThrProThrThrSerGlySerThrValThr..... 469
916 ATAGAAGCCGTCAGCAATATCTTTATGCGAGCATCCCATCAAGGAT 965
    |||||
470 .....ProSerSerLe.....ProGlyThrThrHisThrA 480
966 TGGAGCTGTCGCGGGAATATACGCTGGGCGGATCAGCGACATCTTG 1015
    |||||
480 lArgValLeuThrThrThrThrThrThrValAlaThrGlySerMetAla 496
1016 TCAGCGGTGCGAGATGGCGCGATCGCATTCGCAAGAGGAATCCGCC 1065
    |||||
497 ThrPro.....SerSerSerThrGlnThrSerGlyThrProPr 509
1066 CTCAGCGACATTTTGGCGATGCGGCAATACG..... 1096
    |||||
509 oSerLeuThrThrThrAlaThrThrThrIleThrAlaThrGlySerThrTr 526
1097 ..CAAAATACCGCGCTCCCTTACCATTCGCGAAATATCGCTCAAC 1144
    |||||
526 snProSerSerThrProGlyThrThrProIleProValLeuThrSer 542
1145 AGCAGCGTTACG.....GCAAGAAATATCACTCTCTCA 1179
    |||||
543 ThrAlaThrThrProAlaAlaThrSerSerLysAlaThrSerSerSe 559
1180 ACCGTGCGCGCGCTCAACGCGCAAAATGCAAAATGCGAGACGACGCA 1229
    |||||
559 rPro.....ArgThrAlaThrThrLeuProValLeuThrSerThrAla 574
1230 CCGGAAGACAGCGCTTACCGTTTGACG 1255
    |||||
574 hLysSerThrAlaThrSerPheThr 582
seq_name: sp_fungi:Q92223
seq_documentation_block:
ID Q92223 PRELIMINARY; PRT; 961 AA.
AC Q92223;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CHITINASE.
GN CHTA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC89;
RX MEDLINE=98162139; PubMed=9501518;
RA Takaya N., Yamazaki D., Horiuchi H., Ohta A., Takagi M.;
RT Cloning and characterization of a chitinase-encoding gene (chla) from
RT Aspergillus nidulans, disruption of a chitinase-encoding gene (chla) from
RT frequency and hyphal growth.
RL Biosci. Biotechnol. Biochem. 62:60-65(1998).
CC -| SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL: D87895; BAA36223.1.
DR HSSP: P23472; 2HW.
DR InterPro: IPR001579; Chitinase_2.
DR Pfam: PF00192; chitinase_2; 2.
DR PROSITE: PS01095; CHITINASE_18; 1.
  
```

KW Glycosidase; Hydrolase.
SO SEQUENCE 961 AA; 96839 MW; 66263B3325C533E2 CRC64;

alignment_scores:
Quality: 152.00 Length: 436
Ratio: 0.724 Gaps: 18
Percent Similarity: 48.165 Percent Identity: 22.936

alignment_block:
US-09-303-518D-463 x Q92223 ..

Align seg 1/1 to: Q92223 from: 1 to: 961

```

97 ATCCGCGCAGTT.....CTCGACGCTGACATTTGACCCGACG 137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
332 ILEARGIILEUTYRASPLEUASPROAANHIS.PROPROPTH 348
|||:|||||:|||||:|||||:|||||:|||||:|||||:
138 GAATACCACTATTGCGACGAGGGGGAGCTTGCCNAGCGCAAGGCC 187
|||:|||||:|||||:|||||:|||||:|||||:|||||:
348 hrserProthProthProth.....ProserThrThrThr 360
|||||:|||||:|||||:|||||:|||||:|||||:
188 ATTCGATGGGAACATCAAGACATCAGTTGGGCCACTGATGANT 237
|||:|||||:|||||:|||||:|||||:|||||:|||||:
361 ThSerThrThrSerThr.....ThrSerThrThrThr 366
|||||:|||||:|||||:|||||:|||||:|||||:
238 CAACAGCGCGCTTGAGGAATATCGCTACATTTGCCGTTTCCGA 287
|||||:|||||:|||||:|||||:|||||:|||||:
367 .....ThSerThrThrSerThrThrThrThrThrThr 375
|||||:|||||:|||||:|||||:|||||:|||||:
288 TCAGCGGCAATTCATTCGCTTGCACCAACATGCTTCACATTCG 337
|||||:|||||:|||||:|||||:|||||:|||||:
375 hrThSerThrThrSerThrThrThrThrThrThrThrThr 386
|||||:|||||:|||||:|||||:|||||:|||||:
338 ATTCGACGAGCGGAGTCCGCTTGCATTCAGCTTACGCGCATC 387
|||||:|||||:|||||:|||||:|||||:|||||:
387 .....ThrThrProthThrThrThrThrThrThrThrThr 398
|||||:|||||:|||||:|||||:|||||:|||||:
388 CATTGGACGAGTACG...AACACATCCCGCGACGCTATGACGGCC 434
|||:|||||:|||||:|||||:|||||:|||||:|||||:
398 rThrThrThrProthProthProthProthProthProthProth 415
|||||:|||||:|||||:|||||:|||||:|||||:
435 ACAGCGGGGGCTATCCGCTCCCAAGCGCGAGGATATACAGCT 484
|||||:|||||:|||||:|||||:|||||:|||||:
415 hrThrGlutThrValThrProthProthProthProthProth 426
|||||:|||||:|||||:|||||:|||||:|||||:
485 ACGACATAAAGCGCTTGCCCAAAATATCCGCTCAACCTGACGCAAC 534
|||||:|||||:|||||:|||||:|||||:|||||:
427 .....ProserGlutSerThrThrSerGlutThrSe 437
|||||:|||||:|||||:|||||:|||||:|||||:
535 CGCAGACCGGACACGGCTTGCCGACCGTTTCCACATGCGCGCTAT 584
|||||:|||||:|||||:|||||:|||||:|||||:
437 rSerLeuProserThrThrSerThrProValSerGlutThrProser 452
|||||:|||||:|||||:|||||:|||||:|||||:
585 GCTGACCAAGAGTAGCGAGGATTCAAAGCGCCACCC 625
|||||:|||||:|||||:|||||:|||||:|||||:
453 .....GlutThrLysThrProthThrSerSerThrAlaProthLeuSer 467
|||||:|||||:|||||:|||||:|||||:|||||:
626 .....GATACAGC 633
|||||:|||||:|||||:|||||:|||||:|||||:
468 SerSerProValGlyGlySerSerThrAlaSerSerThrThrThr 484
|||||:|||||:|||||:|||||:|||||:|||||:
484 rProserGlutThrProserAlaSerSerThrAlaValSerGlutThr 501
|||||:|||||:|||||:|||||:|||||:|||||:
678 CACTGCAGATATGCTCAAAACATCATCGCGCGGAGGAGGAATG... 724
|||||:|||||:|||||:|||||:|||||:|||||:
501 ertThrHisLysSerThrThrSerSerGlyProLutThrSerLeuThr 517
|||||:|||||:|||||:|||||:|||||:|||||:
725 ...TCGCGCAGCGGATCCGTCAGGAGGTAAAGCGAAGGCTCAAACT 771
|||||:|||||:|||||:|||||:|||||:|||||:
518 GlySerSerThrSerValProAlaThrSerSerSerValProSerSer 544
|||||:|||||:|||||:|||||:|||||:|||||:

```

```

772 GCTGTCAATGACGCGTTGGGTCTGCTTT.....CCACCGA 806
|||||:|||||:|||||:|||||:|||||:|||||:
534 alSerProserSerThrProValIleSerGlutThrProAlaProPro 551
|||||:|||||:|||||:|||||:|||||:|||||:
807 AAACAAGATGGCGCGCATCAAGCATTTGGCAGATATGGCCAGCTCAAG 856
|||||:|||||:|||||:|||||:|||||:|||||:
551 alThrSerSerSerSerThrThrPhenAlaSerSerThrThrThr 567
|||||:|||||:|||||:|||||:|||||:|||||:
857 ACTATGCCGACGACGACATCCGCGATTTGGCAG..... 889
|||||:|||||:|||||:|||||:|||||:|||||:
568 AspysSerGlutSerSerThrAlaIleGlyThrHisSerSerSeril 584
|||||:|||||:|||||:|||||:|||||:|||||:
890 .TCCAAACCCCAATGCCGCAAGCATAGAACCGCTGACGCAATTCCT 938
|||||:|||||:|||||:|||||:|||||:|||||:
584 eSerGlutThrPro.....SerAlaSerThrProAlaAlaSerPro 598
|||||:|||||:|||||:|||||:|||||:|||||:
939 TATGCGACCATCCCATCAAGGATTTGGACCTG..... 973
|||||:|||||:|||||:|||||:|||||:|||||:
598 ertThrSerProLutThrThrLysThrLeuThrValPheProthProGly 614
|||||:|||||:|||||:|||||:|||||:|||||:
974 TCCGCGGAAATACGCGCTTGCGCGCATCACGGCACATCTGTCAAGCG 1023
|||||:|||||:|||||:|||||:|||||:|||||:
615 SerSerValSerThrGlyThrThrSerAlaSerThrLeuSerSerVal 631
|||||:|||||:|||||:|||||:|||||:|||||:
1024 TCGCAGATGGCGCGCATTCGCTCCGAAGGAAATCCCGCGTACGCA 1073
|||||:|||||:|||||:|||||:|||||:|||||:
631 lProAlaThrSerSerGlyHisThrGlutThrSerThrValSerThr 648
|||||:|||||:|||||:|||||:|||||:|||||:
1074 CAATTTGCCGATCGGCATACGCCCAATACCCGCTCCCTTACATTC 1123
|||||:|||||:|||||:|||||:|||||:|||||:
648 ertAlaSerGlutThrProserAlaSerThrSerLysProLeu... 663
|||||:|||||:|||||:|||||:|||||:|||||:
1124 GAATATCCGTTCAACTGTGAGC.....AGCGTTACGCGCAAGAAAC 1167
|||||:|||||:|||||:|||||:|||||:|||||:
664 ThrsSerAlaSerSerThrThrGlySerValThrThrProse 680
|||||:|||||:|||||:|||||:|||||:|||||:
1168 ATCACTCTCTCAACCGTGGCGCGCT.....CAACGCAAAA 1205
|||||:|||||:|||||:|||||:|||||:|||||:
680 rAlaProGlyValProserSerSerAlaGlySerAspLutThrAlaThr 697
|||||:|||||:|||||:|||||:|||||:|||||:
1206 TGTCAATGCGACAGCAACGCCACCGCAAGCAGGCTACCGTTGACG 1255
|||||:|||||:|||||:|||||:|||||:|||||:
697 hrSerThrThrAspSerGlutProthSerThrSerSerGlySerValThr 713
|||||:|||||:|||||:|||||:|||||:|||||:
1256 GTAAAG 1261
|||||:|||||:|||||:|||||:|||||:|||||:
714 Allys 715
|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: sp_rodent:070495
seq_documentation_block:
ID 070495 PRELIMINARY; PRT; 897 AA.
AC 070495;
DT 01-AUG-1998 (TEMBLrel. 07, Created)
DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE PLENTY-OF-PROLINES-101.
GN SRMT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN.
RA Vayssi re B.M., Camonis J.H.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF062655; Acc17422.1;
DR MGD; MGI:1858303; Strml.
DR InterPro; IPR002483; PWT.
DR Pfam; PF01480; PWT.1.
DR SMART; SMO0311; PWT.1.

```

50 SEQUENCE 897 AA; 101166 MW; 62160BB6772BD10 CRC64;

alignment_scores:

Quality: 150.50 Length: 479
Ratio: 0.727 Gaps: 25
Percent Similarity: 43.215 Percent Identity: 23.591

alignment_block:

US-09-303-518D-463 x 070495 ..

Align seg 1/1 to: 070495 from: 1 to: 897

```

84 AAGCATCCCTTTATCCGAGAGTTCTGACCGCTGACGATTTGCAACCG 133
   ::::: ||| ::::: |||
167 GUAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 183
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
134 AGGGAATATACCATTTGCGAGAGGAGGAGGAGGAGGAGGAGGAGG 183
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
183 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 200
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
184 GGCATATCGGATTTGGAAACATACAAAGCATGAGTTGGGCGACCTGAT 233
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
200 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 216
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
234 GATTCACAGGCGCGCG...TTGAAGAAATATGCGCTACATGCTCGCT 280
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 Lys..GluLysSerProGluLeuProGluProSerValArgMetLys... 231
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
281 TTTCGATCAGCGGCGCAAAATTCATTCGCGCTTCGACACCATGCTCA 330
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 .....AspSerSerValGlnGlnLysLysLysLysLysLysLys 243
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
331 CATTCGATTCGATTCGAGAGCGG...GTATCCCGCTGACGAGATTCACCT 377
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
243 PileuLysLysLysLysLysLysLysLysLysLysLysLysLysLys 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
378 TTACCG.....CATCCATTGGG 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 SerProGluLysAsnSerLysLysGluLysLysLysLysLysLysLys 276
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
395 ACGGATACGACCATCCCGCGCGCGCTATGACGCGCGCGCGCGCGCG 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 rArgSerLysSerLysSerLysSerLysLysLysLysLysLysLys 292
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
445 GGCATCCCGCTCCCAAGCGCGCGGATATATACAGTACGACATPAA 494
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
292 erHisThrArgProArgArgArgHis.....ArgSerArgSerArg 305
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
495 AGGCGTTGCCCAAAATATCCGCTCAACCTGACCGACACCGCGAG...CA 541
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
306 SerLysSerProArgArgArgProSerProArgArgArgProSerPro 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
542 CCGGACAAACGCTTCGCGCGCTTCCCAATGCGCGCGCTATGCTGACG 591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
322 gArgArgThrProArgArgArgMetProProProProArgHisArgArg 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
592 CAAGGAGT.....AGCGGAGCAT..... 610
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339 erArgSerProGluArgArgArgArgArgSerSerAlaSerLeuSerGly 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
611 .....TCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
356 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 372
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
634 CCCAGCTGACAGATCGGCAATGCGCGCGCGCGCGCGCGCGCGCGCG 683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
372 oProLysArgThr.....SerSerProProGluLysLysLysLysLys 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
684 AGATATGTCATAAAACATCATCGCGCGCGCGCGCGCGCGCGCGCGCG 733
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
387 er...ProSerAlaSerProProArgArgArgArgHisArgProSerSer 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

734 GCGATGCCGTCGACGGGTATAGCGAAGCTCAAAACATTTGCTCATGCAC 783
   ||| |||
403 AlaThrPro..... 405
784 GCGTTGGGCTGCTTTCCACGAAACAGATGGCGGCATCAACGATTT 833
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
406 .....ProProLysThrArg...HisSerProThrProG 416
834 GGCAGATATGCCGCACTCAAGACT.....ATCGCGACGACG 871
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416 InGlnSerAsnArgThrArgLysSerArgValSerValSerProGluArg 432
872 CCATCCGCG.....ATTGGCATTCCAAAACCCCAATGCC 906
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
433 ThrSerGluLysValThrLysHisLysGlyThrGluLysArgLysSerP 449
907 GCACAAAGCATAGAACCGCTCAGCAATATCTTATGCGACCATTCCTCC 956
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
449 oSerProAlaProLysProArgLysValGluLeuSerGluSerGluGln 466
957 CAAGGAGATTGAGCTGTCGCGGGAATATACGCTTGGCGCGCATCAGG 1006
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
466 sPylsGlySerLysMetAlaAlaAlaAspSerValGlnInArgArgLys 482
1007 CAC.....ATCCT 1014
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 TyrArgArgLysAsnGlnInGlnSerSerSerAspSerGlySerSerTh 499
1015 GTCACACGCTGCGCAGATGGCGCGCATTCGATTCGCAAGAGCAAT... 1060
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
499 rSerGluAspGluArgProLysArgSerHisValLysAsnGlyGluValG 516
1061 .....CCGCGCTCAGCAGCAATTTGCCGATGCGGCAT 1093
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
516 LysArgArgArgArgLeuSerProSerArgSerAlaSerPro..... 529
1094 ACGCAAAATACCGTCCCTT...ACCATCCCGAAATATCCGTT...CA 1137
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
530 SerProArgLysArgGlnLysGluThrSerProAlaGlnMetGlnLys 546
1138 AACTTGACGACGCTTACGCGCAAAAGCA..... 1168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
546 sArgTrpGlnSerProValThrLysSerSerArgArgArgSerPro 563
1168 ..... 1168
563 erProProProAlaArgArgArgArgSerProSerProAlaProProPro 579
1169 ...TCACCTCTCAACCGTCCGCGCTCAACGCGCAAAATATGCAACGTG 1215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
580 ProProProProProProProProArgArgArgArgArg..... 591
1216 GCAGACCAACGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
592 .....ProThrProProProArgArgArgThr 600

```

seq_name: sp_human:09036

seq_documentation_block:

ID 09036 PRELIMINARY; PRT; 1275 AA.

AC 09036;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE RNA BINDING PROTEIN (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Ontaki S., Umeki K., Sawada Y.;

RT "Homo sapiens mRNA for RNA binding protein, partial cds.";

RU Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016091; BAA83717.1; -
 FT MON_TER 1
 SQ SEQUENCE 1275 AA; 136869 MW; 45C2B2F85E98A6F6 CRC64;

alignment_scores:
 Quality: 150.00 Length: 598
 Ratio: 0.558 Gaps: 28
 Percent Similarity: 44.983 Percent Identity: 22.742

alignment_block:
 US-09-303-518D-463 x Q9UQ36 ..

Align seg 1/1 to: Q9UQ36 from: 1 to: 1275

```

39 ACTGGCAGTGTGCT...GCCGATGCATGCACACGCCCTCAGATTGGCAA 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
102 SerGlySerSerProGluValAspSerLysSerArgLysSer...ProAr 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86 ACATCCCTTATCCGGCAGTTCGACCCGTCGACATTCGACACCCGAC 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 gArgSer...ArgSerGlySerSerProGluValLysAspLysProA 132
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 GGGAAATACCACTATTCGGCAGCAGGGGAGCTTCCNAGCGCAACG 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 rGalaIaPro...ArgAlaGlnSerGlySerSpSerSerProGlu 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 CCATATCGATTGGGAACATACAAAGCCATCATCTGGGCCACCTGATGA 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
147 ProLysAlaProAlaProAlaLeuProAlaArgArgSerArgSerGly 163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 TTCACACGGGGCCGTTGAAGAAATATGGCTACATTCGCGCTTTTCC 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
163 rSerSerLysGlyArg...GlyProSerProG 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 GATCAGCGGCGACAAATTCATTCGCCCTTCGACACACATGCTCACATTC 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
173 LueGlySerSerSerThrGlnSerSerProGluHisProPro... 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 CGATTTCAGCAAGCCGCTAGTCCCGTTGACGATTACGCTTACCAGA 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
187 .....LysSerArgThrAlaArgArgGlySerArgSerSerProG 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 TCCTATGGGCGGATACGACACCATCCGCGGAGCTGACGGGCGCA 435
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 uProLysThrLysSerArgThrProProAlaArgArgSerSerArgS 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
436 CAGGCGCG...CGGCTATCCGCTCCCAAGG... 464
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 erProGluLeuThrArgLysAlaArgLeuSerArgArgSerArgS 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
465 .....CGGAGGATATATACAGTACGACATCAAAAGCGC 499
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 SerSerSerProGluThrArgSerArgThrProProAlaHisArg 250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
500 TTGCCCAAAATATCCGCTACACCTACGACAA... 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 rProSerValSerSerProGluProAlaGlnLysSerArgSerSerArg 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
534 .....CCGACGACCGGACACAGCGCTTCCGCA... 560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
267 rGArgArgSerAlaSerSerProArgThrLysThrThrSerArgArg 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
561 .....CGTTTCACAAATGCGGCGCTATGCTGACGACAGAGTAG... 602
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
284 ArgSerProSerProLysProArgGlyLeuGlnArgSerArgSerArg 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
603 .CGACGATTCAAAAGCGCCACCG...ATACAGCCCGGAGCTGACAGA 648
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 rArgArgGlnLysThrArgThrThrArgArgArgAspArgSerGly... 315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
649 TCGGCAATGCCGCGGAAGCTTCAAGCGCAGCTCAGATATCGTCAAAA 698

```

```

316 .....SerSerGlnSerThrSerArgArgGlnArg 326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
699 CATCATCGGCGCGGACAGAGAAATTCGCGGACGAGCGCTGACG 748
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
327 SerArgSerArgSerArgValThr...ArgArgArgArgGlyLysSer 342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
749 GTATACCGAAGGCTCAACATTCGCTGCATGCACGCGCTTGCTGCT 798
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 YThr...HisSerArg... 346
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
799 TCACCGCAAAACAAAGATGCGCGCATCAAGATTGGCAATATGGCG 848
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 .....SerProAlaArgGlnGlnSerSerArgThrSerSer 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
849 ACTCAAGACATATGCGCAGACGACATCCGATTCGGACGCCAAAC 898
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
359 ArgArgArgArgArgGlyArgSerArgThrPro... 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
899 CCAATGCGCGCACAAGGCAATAGACCGCTCAGCAATATCTTATGCA 948
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369 oThrSerArgLysArgSerArgSerArg...ThrS 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
949 ATCCC...CATCAAGGATTCGACCTGCTCCGGGGAATA 986
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
380 erProAlaProThrLysArgSerArgSerArgAlaSerProAlaThr 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
987 CGGCTGGGCGGCATACAGCGCATCTCTGTCAGCGGTCGAGATGGGC 1036
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397 ArgArgSerArg...SerArgThrProLeuLysSerArgArg 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1037 CGATGCAATTCGCAAGGAATCCGCGCTCAGCGCACATTTTCCGAT 1086
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
410 gSerArgSerArgThrSerProValSerArgArgArgSerArgSerArg 427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1087 GCGCATACGCCAATATACCCCTCCCTTACCATTCGCAATATCCGTC 1136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
427 hrSerVal... 429
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1137 AAACCTTGAGACAGCGTTACGGCAAAAGAAACATCACTCTCAAC 1186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
430 .....ThrArgArgArgSerArgSerArgAlaSerProVal 441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1187 CGCGCTCAAGCGCAAAATGTCAAACTGCGCAGACCAACGCCACCG 1236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
441 lSerArgArgArgSerArgSerArg...ThrProValT 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1237 ACAGGCGTACCGTTGACGCTAAAGCGTTTCCGAATTTGAGAAGCA 1286
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 hrArgArgArgSer...ArgSerArgThrProThrThrArgArg 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
469 rArgSerArgThrProProValThrArgArgArgSerArgSerArg 486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1337 CTAAAGCTAAGCCTGTGTTGATGCGAAACGAGATGGGAGTGTATG 1386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
486 ro...ProValThrArgArgArgSerArg...Ser 495
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1387 AAGCTTAATTAATTCACAACCTGAGCAGGTGAGAAATATGTTACA 1436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
496 ArgThrSerProIleThrArgArgArgSerArgSerArgThrSerPro 512
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1437 AAGCAGAAAGAGAGTCAAGTAGTCAAGTTAAAGCCATGCGCAAG 1486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
512 lThrArgArgArgSerArgSerArgThrSerProValThrArgArg 528
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1487 AATGGAAATATAAACAGGTTAGATTTAATCATTTATAGGTGTATG 1536
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
529 .....Ser 529
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1537 ATCAATTAAGAAAGCAGATACAGAGAGGACATAGTCAATCCGTGTGA 1586
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

935 hSerial..... 937
1137 AACCTTGAGCAGCTTACGCGAAGAAATCACTCCCTCACCGTGC 1186
938 .....ThraArgArgSerArgSerArgAlaSerProva 949
1187 CCGCCGCAAGCGCAAAATGTCACAGTCGACGACCGCCACCGAAG 1236
949 lSerArgArgSerArgSerArg.....ThraProProvalT 962
1237 ACAGCGCTACCGCTTGACGCGTAAGGGTTCCGAATTTTGAGACACGT 1286
962 hrArgArgArgSer...ArgSerArgThr.ProThrThraArgArgSe 977
1287 GAAATATGATACGACGCTCGATATTCAGATATTCGGGGCGGTATAC 1336
977 rArgSerArgThrProProvalThraArgArgSerArgSerArgThrp 994
1337 CTAAAGCTAAAGCCTGTGTGATGCGAAMCGAGATGGAGGTGTATAG 1386
994 ro.....ProvalThraArgArgSerArg.....Ser 1003
1387 AACCTTAATTAATTTGACAACTCGTGACAGCTGAGAAAATGTTACAGA 1436
1004 ArgThSerProThrThraArgArgSerArgSerArgThSerProva 1020
1437 AACGAGAGAGAGATCAGACAGTACGATTAAAGCCCATGGCGCAACAG 1486
1020 lThraArgArgSerArgSerArgThSerProvalThraArgArgArg. 1036
1487 AATGGGAAATAAACAGGTTAGATTATTAATCATTTTATAGGTGTGAT 1536
1037 .....Ser 1037
1537 ATCAATAGAAAGCAGACAGTACAGAGAGGCGATCTAACCCGTGTGA 1586
1038 ArgSerArgThSerProvalThraArgArgSerArgSerArgThrp 1054
1587 TGTACGGGTATACAAACAACTCGCGCA.....CCGANA 1621
1054 oProAlaAlaArgArgArgSerArgThrpThraProleuThraArg 1071
1622 AACATGGGGTTTATCAAGCAGCAGCATGGAATTAAGCCCTGATGGAAGT 1671
1071 ysaArgSerArgSerArgSerProleuAlaAlaArgArgArgSerArgSer 1087
1672 TGGGAGGTGAAGAAAGGTGGGAAGATGATACCAAG 1713
1088 ArgThraArgThraAlaArgGlyLysArgSerLeuThraArg 1101
seq_name: sp_human:060382
seq_documentation_block:
ID 060382; PRELIMINARY; PRT: 1791 AA.
AC 060382;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE KIAA0324 (FRAGMENT).
GN KIAA0324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
OX
RN
RP SEQUENCE FROM N.A.
RA Rieke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA Goodwin L., Bryant J., Tesmer J., Melnick L., Longmire J., White S.,
RA Deng S., Tatum O., Campbell C., Fawcett J., Deaven L.,
RT "Sequencing of Human Chromosome 16p13.3".
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA Rieke D.O.;
RT "Large Scale Sequence Analysis and Annotation with the Sequence
RT Comparison Analysis (SCAN) System."
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC004453; AAC08453.1; -.
FT NON_TER 1
FT SEQUENCE 1791 AA; 191306 MW; 3A7B5530AEB95F3E CRC64;

```

```

alignment_scores:
  Quality: 150.00 Length: 598
  Ratio: 0.558 Gaps: 28
Percent Similarity: 44.983 Percent Identity: 22.742

```

```

alignment_block:
US-09-303-518D-463 x 060382 ..

```

```

Align seg 1/1 to: 060382 from: 1 to: 1791

```

```

39 ACTGACAGTGTGCT...GCCGATGATGACACAGCCTCAGATTGGCAA 85
611 SerLysSerProGluValAspSerLysSerArgLeuSer...ProAr 626
86 ACGATCCCTTTATCCGCGAGGTTCTCGACCGTCAGCATTTGCAACCCGAC 135
626 gArgSer.....ArgSerGlySerSerProGluValAspLysProA 641
136 GGGAAATACCACTTATTCGCGACGAGGGGAGGCTGCCNAGCCACAGG 185
641 rGaAlaAlaPro.....ArgAlaAlaInSerGlySerAspSerProGlu 655
186 CCATATCGATTGGGAAACATACAAACCATCACTGGGCCACCTGATGA 235
656 ProLysAlaProAlaProAlaProAlaLeuProAlaArgSerArgSerGlyse 672
236 TTCACACAGCGCGCGTTGAAGAAATATCGCTACATTTCCGCTTTC 285
672 rSerSerLysGlyArg.....GlyProSerProG 682
286 GATCAGCGGACAAATTCATTCGCTTCGACAAACCATGCTCACAATTC 335
682 lGlySerSerSerThraLysSerSerProGluHisProPro..... 695
336 CGATTCTGACGAGCGGCTAGTCCGTTGACGATTCAGCTTTACCGCA 385
696 .....LysSerArgThraAlaArgGlySerArgSerSerProG 709
386 TCCATTGGGAGGATACGACACACCATCCCGCGAGCGGTATGACGGGCA 435
709 uProLysThraLysSerArgThraProAlaArgArgArgSerArgSer 726
436 CAGGCGCG.....CGGCTATCCCGCTCCCAAGG..... 464
726 erProGluLeuThraArgLysAlaAlaGluSerArgSerArgSerAla 742
465 .....CGGAGGAGATATATACAGTACGACATTAAGGCG 499
743 SerSerSerProGluThraArgSerArgThraProAlaGluHisArgArgSe 759
500 TTGCCCAAAATATCCGCTCAACCTGACCGCA..... 533
534 .....CCGACGACCGGACCAACGGCTTGCCGA... 560
776 rGArgArgSerAlaSerSerProArgThraLysThraThraSerArgGly 792
561 .....CCGTTTCACAAATGCGCGCGCTATGCTACGCAAGAGAGAG.. 602
793 ArgSerProSerProLysProArgGlyLeuGlnArgSerArgSerArgSe 809

```



```

603 .CGACGATTCGAAACGCCGCCACCCG...ATACAGCCCGAGCTGACAGA 648
      ||||| :|||
809 rArgArgIglulThrArgArgArgArgSerIy... 824
      ||||| :|||
649 TCGGCGATTCGCCCGAAGCTTCAAGCGCAGCTGACATATCGTCAAAA 698
      ||||| :|||
825 .....SerSerInSerThrSerArgArgIglArg 835
      ||||| :|||
699 CATCATGCGCGCGGAGAGAAATTCGGCGGAGGATCGCGTACG 748
      :||| :|||
836 SerArgSerArgSerArgValThr..ArgArgArgArgIglArg 851
      ||||| :|||
749 GTATACGGAAGGCTCAAACTGCTGATGACGCGCTGGTCTGCTT 798
      ||||| :|||
851 YThr.....HisSerArg..... 855
      ||||| :|||
799 TCCACCGAAGAACAGATGGCGGATGACGATTTGGCACAATATGGCGCA 848
      :||| :|||
856 .....SerProIlaArgIglInSerSerArgThrSerSer 867
      ||||| :|||
849 ACTCAAGACTATGCGCAGCAGCCATCCGATGGGAGCTCAAAAC 898
      :||| :|||
868 ArgArgArgArgIglArgSerArgThrPro.....Pr 878
      ||||| :|||
899 CCATGCGCGCAGCAGCATGAAACCGCTGACGATCTTATGCGAGCC 948
      ||||| :|||
878 oThrSerArgIglArgSerArgSerArg.....Thrs 889
      ||||| :|||
949 ATCC.....CATCAAGGATGTGAGCTGCGGCGAATA 966
      ||||| :|||
889 erProIlaProThrPlyArgSerArgSerArgIlaSerProIlaThrHis 905
      ||||| :|||
987 CGGCTGGCGGCGATCATCGCACATCTGTCAAGCGGTGCGAGATGGCG 1036
      ||||| :|||
906 ArgArgSerArg.....SerArgThrProIleuIleSerArgArg 919
      ||||| :|||
1037 CGATTCGATTCGCGAAGGAAATCCCGCTGACGCGCAATTTGCCGAT 1086
      ||||| :|||
919 gSerArgSerArgThrSerProValThrArgArgArgSerArg 936
      ||||| :|||
1087 GCGGCATACGCCAATACCCGCTCCCTTACCATTCGCGAATTCGCTTC 1136
      ||||| :|||
936 hrSerVal..... 938
      ||||| :|||
1137 AAACCTGGAGCAGCTTACGGCAAGAAACATCACCTCTCAACCGTGC 1186
      ||||| :|||
939 .....ThrArgArgSerArgSerArgIlaSerProVa 950
      ||||| :|||
1187 CGCGCTCAACGCGCAAAATGTCAACTGCGACAGCAAGCCACCGAAG 1236
      ||||| :|||
950 lSerArgArgArgSerArgSerArg.....ThrProValT 963
      ||||| :|||
1237 ACAGCGGTACCGCTTTCAGCGTAAGGTTCCGATTTTGAGAGCAGCT 1286
      ||||| :|||
963 hrArgArgArgSer...ArgSerArgThr.ProThrThrArgArgArg 978
      ||||| :|||
1287 GAATATGATACGAGCTGATATTCAGAAATTAATGGGGGCGGTATAC 1336
      ||||| :|||
978 rArgSerArgThrProProValThrArgArgArgSerArgThrP 995
      ||||| :|||
1337 CTAGAGCTAGCGCTGTGTTGATGCGAACCAGATGGAGGTATAGG 1386
      ||||| :|||
995 ro.....ProValThrArgArgSerArg.....Ser 1004
      ||||| :|||
1387 AACCTTAATTAATTCAGAACTGTCGACGAGTGAAGAAATGTCAGGA 1436
      ||||| :|||
1005 ArgThrSerProIleThrArgArgArgSerArgThrSerProVa 1021
      ||||| :|||
1437 AACGAAAGAGAGAGTACAGTAGTCACTTAAGCCCATGCGCAACAG 1486
      ||||| :|||
1021 lThrArgArgArgSerArgSerArgThrSerProValThrArgArgArg 1037
      ||||| :|||
1487 AATGGGAATAAAGAGGTAGATTTAATCATTTATAGGTGTAT 1536

```

```

1038 .....Ser 1038
1537 ATCAATAGAAAGCAGACTAACAGAGGCGATCTAACCCTGGTGA 1586
      :||| :|||
1039 ArgSerArgThrSerProValThrArgArgArgSerArgThr 1055
      ||||| :|||
1587 TGTACGGTGTATACAAACACTCGCA.....CCTGATA 1621
      :||| :|||
1055 oProIlaIleArgArgSerArgSerArgThrProIleuProArgL 1072
      ||||| :|||
1622 AACATGGCTTATATCAAGCAGACGAGAAATTAAGGCTCATGAGT 1671
      ||||| :|||
1072 ysArgSerArgSerArgSerProIleuAlaIleArgArgArgSer 1088
      ||||| :|||
1672 TGGAGGTGAAGCAAGAAAGGTGGGAAGTATGACCAAG 1713
      ||||| :|||
1089 ArgThrProArgThrAlaArgIglArgSerSerLeuThrArg 1102
      ||||| :|||
seq_name: sp_human:043418

```

```

seq_documentation_block:
ID 043418 PRELIMINARY; PRT; 513 AA.
AC 043418;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE INTESTINAL MUCIN (FRAGMENT).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97476275; Pubmed=9334251;
RA Gum J.R. Jr., Ho J.J.L., Pratt W.S., Hicks J.W., Hill A.S.,
RA Vinal L.E., Robertson A.M., Swallow D.M., Kim Y.S.;
RT "MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl
RT terminus, and a novel upstream repetitive region."
RL J. Biol. Chem. 272:26578-2666(1997).
DR EMBL; AF007190; AAC02268.1; -.
FT NON_TER 1
FT NON_TER 513
FT SEQUENCE 513 AA; 52573 MW; 33BBAC2AFABE1436 CRC64;

```

```

alignment_scores:
Quality: 149.00 Length: 485
Ratio: 0.734 Gaps: 19
Percent Similarity: 41.856 Percent Identity: 20.825

```

```

alignment_block:
US-09-303-518D-463 x 043418 ..

```

```

Align seg 1/1 to: 043418 from: 1 to: 513
62 ATGCACAGCTCGAGATTGGCAACGATCCCTTATCC..... 100
      ||||| :|||
52 MetIleThrSerHisThrThrAsnThrThrProIleuSerThrLeuValTh 68
      ||||| :|||
101 .GGCAGGTTCTGCAGCGTCAGCATTTGCAACCGCGGAAATACGACCT 149
      ||||| :|||
68 rThrLeuThrThrThrIleThrArgSerThrProThrSerIglThrThrT 85
      ||||| :|||
150 ATTCGGAGCAGAGGGGAGGAGCTTG.....CCMAGCGCAGGCGCATATCG 193
      ||||| :|||
85 YrProThrSerProThrSerIleValSerAspSerThrThrIguille... 100
      ||||| :|||
194 GATTGGGAACATACAAA.....GCATCAGTTGGGCGCAC 228
      ||||| :|||
101 .....ThrYrSerThrSerIleThrIglYThrLeuSerThrAlaThr 114

```

```

229 CTGATGATTACACAGCGCGCGTGAAGAAATATCGGCTACATTGCGG 278
114 r.....ThrLeuProProThrsers 122
279 CTTTCCGATCAGCGGACAAATTCATCCGCTTCGACAAACATG... 325
122 erLeuProThrThrGluThrAlaThrMetThrProThrThrLeuile 138
326 .....CCTCAC.....ATCCGATTCGACGAAAGCC 351
139 ThrThrProAsnThrThrsHisSerThrProSerPheThrSerSe 155
352 GGTATCCCGTTCGAGGATTCAGCCTTACC..... 382
155 rThrIleYSerThrValSerThrThrAlaIleSerSerAlas 172
383 .....GCATCC 388
172 erProThrseryThrMetValThrSerThrThrMetThrProSerSer 188
389 ATTGGAGCGATACGACACCATCCCGCGGAGGCTATG..... 427
189 LeuSerThrAspThrProSerThrThrProThrIleThrTyProse 205
428 .....ACGGGCGACAGCGCGGCTATCCGCTC..... 457
205 rValGlySerThrGlyPheLeuThrAlaThrAspLeuThrSerThrp 222
458 .....CCAAAGCGG 466
222 heThrValSerSerSerAlaMetSerlySerValIleProSerSer 238
467 CGAGGATATATACAGCTACGACATAAAGCGTTCGCCAAATATCCG 516
239 ProSerIleGlnAsnThrGluThrSerLeuValSerMetThrSerAl 255
517 CTCACAC.....TGACGACAAACGCGACACCGGAGCAAGCGCTTG. 556
255 aThrThrProSerLeuArgProThrIleThrSerThrAspSerThrLeu 272
556 ..... 556
272 hrSerSerLeuThrThrPheProSerThrTySerPheSerSerSer 288
557 .....CCGACCGTTTCACAAAT 573
289 MeSerAlaSerSerAlaGlyThrThrHisThrGluThrIleSerSerle 305
574 GCGGCGGCTATGCTGACGACAGAGTAGCGGAGATTCAACGCGCCAC 623
305 uproAlaSerThrAsnThrIleHisThrAlaGluSerAlaLeuAlaP 322
624 CGCATACAGCCCGAGCTGGACGATCGGCGCAATGCCGCGGAAGCTTCA 673
322 roThrThrThrThrSerPheThrThrSerProThrMetIuProProSer 338
674 ACGGCACTGCAGATATCGTCAAAAACATCATCGCGCGGAGAGAAATT 723
339 ThrThrValAlaThrThrGly..... 345
724 GTGCGCGCAGGCGATCGGTCAGGAGTAAAGGAAGCTCAAAACATTGC 773
345 ..... 345
774 TGTGTCACAGCGCTGGTCTGCTTCACCGAAACAGATGGCGCGCA 823
346 .....ThrGlyGlnThrThrPheProSerSerThr..... 355
824 TCAACGATTTGGCAGATAGCGCAACCTCAAGACGTATGCCGAGACGCC 873
356 AlaThrPheLeuGluThrThrThrLeuThrProThrThrAspPheSerTh 372
874 ATCGGCGATTGGGCGATCCCAAAACCCCAATGCCGCAAGGATAGAGAG 923

```

```

372 rgIuSerLeuThrThrAlaMetThrSerThrp.....P 384
924 CGTCACCATATATCTTATATGACGACCATCCCATCAAGAGATTGGAGCTG 973
384 rollerThrSerSerle.....ThrProThrAspThrMetThrSer 397
974 TCCGGGAAATATACGGCTTGGCGGATCAGCGACATCCGTCAGAGCGG 1023
398 MeTarGthrThrThrSerThrProThrAlaThrAsnThrLeuSerPro.. 413
1024 TCCGACATGGGCGCATTCGCATTCGCAAGGAAATCCGCGTCAGCA 1073
414 .....LeuThrSerSerIleLeuSerSerThrProValProSerT 427
1074 CAATTTCGCGGATGCGGATACGCCAAATACCGGCTCCCTT..... 1114
427 hrGluValThrThrSerThrAsnThrAsnThrProValSerThrLeu 443
1115 .....ACCATCC...GAATATCCGTTCAAACTTGAGAGAGCGTTAC 1155
444 ValThrThrLeuThrProThrIleThrArgSerThrLeuThrSerGluTh 460
1156 GCGAAGAAACATACCTCTCTCAACCGTCCGCGCTCAAGCGCAAAA 1205
460 rAlaTyProSerSerProThrSerThrValThrGluSerThrThrGluI 477
1206 TGACA 1210
477 leThr 478

seq_name: sp_human:Q9NYE4
seq_documentation_block:
ID Q9NYE4 PRELIMINARY; PRT; 716 AA.
AC Q9NYE4:
DT 01-OCT-2000 (TREMBLrel. 15' Created)
DT 01-OCT-2000 (TREMBLrel. 15', last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15', last annotation update)
DE CERVICAL MUCIN MUC5B (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ho S.B., Toribara N.W., Anway R.E., Spurr-Michaud S.J., Shekels L.L.,
RA Keulman H.T., Hill J.A., Gipson I.K.;
RT "Expression cloning of human cervical proteins using an antibody to
RT cervical mucus."
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253321; AAF64523.1; -.
FT NON_TER 1 1
FT NON_TER 716 716
SQ SEQUENCE 716 AA; 71194 MW; 6F3FE20E7512289F3 CRC64;

```

```

alignment_scores:
  Quality: 148.00      Length: 436
  Ratio: 0.736        Gaps: 24
Percent Similarity: 46.101 Percent Identity: 23.394

```

```

alignment_block:
US-09-303-518D-463 x Q9NYE4 ..
Align seg 1/1 to: Q9NYE4 from: 1 to: 716

```

```

26 TTATCTGTCATAGTGGAGTGTGCTGCGAGATGACAGCGCTCA 75
311 ValPheCys.....CysAsnTyrlGlnHisCysPro.. 320
76 GATTTCGCAAGAGATCCCTTATCGGCGAGGTTCGACCGTCAGCATTT 125

```

```

321 .....S 321
126 CGAACCCGACGGAAATACACCTATTCGAGAGGGAGCTCCN 175
|||||.....
321 erThrAlaThrSerSerThr.....AlaMetPro 331
176 AGCGCAAGCGCCATTCGATTCGGAAACATACAAAGCATCAGTTGGC 225
|||||.....
332 SerSerThrProGlyThrThrTrp..... 339
226 CACCTGATGATTCACAGCGCGCTTGAAAGAAATATGCGTACATTGT 275
|||||.....
340 .....IleuThrGluLeuThrThrAlaThr 349
276 CGCGTTTTCGATCAGCGGACAAATTCATTCGCGCTTGACACA..... 319
|||||.....
349 hrThrThrAlaSerThrGlySerThrAlaThrProSerSerThrProGly 365
|||||.....
320 ACATGCGTCACATTCGATTCGACGAGACCGGTAGTCCGTTGACGGA 369
|||||.....
366 ThrAlaProProProlValLeuThrSerProAla.....ThrThr 379
370 TTCAGCCTTACCGCATCCATTCGAGGAGATACGACATCCCGCGGA 419
|||||.....
379 rProThrAlaThrSerSerThrAlaThrSerSerSer.....ProArgT 395
420 CGCGTATGACGGGCGACAGGGCGGCTATCCCGTCCCAAGGCGCGA 469
|||||.....
395 hrAlaThrThr.....LeuProValLeuThr 403
470 GGGATATATACAGCTACGACATAAAGCGTTGCCAAATATCCGCTC 519
|||||.....
404 SerThrAlaThrLysSerThrAlaThrSerThrAlaThrProLysProSer 420
|||||.....
520 AACCTGACGCAACACCGACACCGGACAAAGCGCTTCCGACCGTTTCCA 569
|||||.....
420 rThrLeuGlyThrThrLeuThr.....LeuProGluGluThrT 433
570 CAATGCGCGGCTATGCTGACGAGAGAGTAGGCGGATTCGAAACGGG 619
|||||.....
433 hrThrProVal.....AlaThrMetSerThrIle 442
620 CCACCGCATACAGCCCGAGCTGACAGATCGGGCAATCGCGGCAAGCC 669
|||||.....
443 HisProSerSerThrProGluThrThr.....HisThr 453
670 TTCAACGGCAGCTGACATATCGTCACAAACATATCGGGCGGACAGAGA 719
|||||.....
453 rSerThrValLeuThrThrGluAlaThrThrThrAlaThrSerSerT 470
720 AATGTGCGGCGAGCGGATGCGG.....TGCAGGGTA.....TAA 754
|||||.....
470 hrSerThrProSerSerThrProGlyThrThrCysIleLeuThrGluLeu 486
755 GCGAAGGCTCAAAATTCGCTGTCATGACGCGCTGGGTGCTTTCACCC 804
|||||.....
487 ThrThrAlaAlaThrThrThrAlaAlaThrGlyProThrAlaMetPro 503
|||||.....
805 GAAACACAGATGCGCGCATCAACATTCGAGCATATGCGGCACTGCA 854
|||||.....
503 rSerThr.....ProGlyThrThrTrp.....IleLeuThrGluLeu 516
855 AGACTATGCGCGGACGACCATCGCGATTCGCGACATCAAAACCCCAATG 904
|||||.....
516 hrThrThrAlaThrThrThrGluSerThrGlySerThrAlaThrPro 531
905 CCGCACAGAGCATAGAGCGCTCAGCAATATCTTATGG.....CA 945
|||||.....
532 .....SerSerProGlyThrThrTyrIleLeuThrGlu 542
946 GCCATCCCATCAAAAGGATTCGAGCTCCGCGGAAATACGAGCTGGG 995
|||||.....
542 uProSerThrThrAlaThrValThrValProThrGlySerThrAlaThr 559

```

```

996 CGGCATACAGG.....CACATCCTGTCAAGCGGTGCG 1027
|||||.....
559 lAsSerThrGluAlaIleGlyThrProHisValSerThrThrAlaThr 575
1028 AGATGGCGCGATCGCATTCGCGA.....AGGCAATCCGCGCTC 1068
|||||.....
576 ThrProThrValThrSerSerThrLysAlaThrProSerSerSerProGly 592
1069 AGGACCAATTTTGGCG...ATCGGCGATAGCGCAATATACCGTCCGCTTA 1115
|||||.....
592 rAlaThrAlaLeuProAlaLeuArgSerThrAlaThrThrProThrAlaT 609
1116 CC.....ATTCGCAAAATATCCGTT...CAAACTTGAGACAGC 1150
|||||.....
609 hrSerPheThrAlaIleProSerSerSerLeuGlyThrThrThrPhrArg 625
1151 GTTAGCGCAAGAAAGAAATCACCCTCTCAACCGTCCGCGCGTCAACGCG 1200
|||||.....
626 LeuSer.....GlnThrThrThrPro.....ThrAl 634
1201 AAAAATGTCAACGTGACAGACCAACGACCGACGAGAGCGGTACCGTT 1250
|||||.....
634 arThrMetSerThrAlaThrProSerSerThrProGluThrValHisThr 651
1251 TGACGGTA 1258
|||||.....
651 erThrVal 653

```

```

seq_name: sp_bacteria:Q924N7
seq_documentation_block:
ID Q924N7 PRELIMINARY; PRT; 1873 AA.
AC Q924N7;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SURFACE PROTEIN PRECURSOR.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MH594;
RX MEDLINE=9081742; Pubmed=9864215;
RA Shankar V., Baghdadyan A.S., Huyke M.M., Lindahl G., Gilmore M.S.;
RT "Infection-derived Enterococcus faecalis strains are enriched in esp,
RL Infect. Immun. 67:193-200(1999).
DR EMBL: AF034779; AAD09858.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
KW signal.
FT SIGNAL.
FT CHAIN. 50 1873 SURFACE PROTEIN
SQ SEQUENCE 1873 AA; 202084 MW; f6094830b980621 CRC64;

```

```

alignment_scores:
Quality: 148.00 Length: 695
Ratio: 0.468 Gaps: 36
Percent Similarity: 45.468 Percent Identity: 21.871

```

alignment_block:

```

US-09-303-518D-463 x Q924N7 ..
Align seg 1/1 to: Q924N7 from: 1 to: 1873

```

```

80 TGGCAAAAGATCCCTTATCCGACGTTCTCG..... 112
|||||.....
604 TrpLysThrGluProAspThrSerAsnValThrAspSerThrLysGlyI 620

```

```

113 ... ACCGTCACATTTGCAACCCGAGCGGAATAATACCATTTGCGCAG 158
620 eValThrValLysIleGlyAsnArgThrPheAspValAspValGluPheA 637
159 CAGGGGGGAGCTTGCCNAGCCGCAACGATTCGATTTGGGAAACATAC 208
637 Ia.....ValLysAlaSerGluAlaMetGluAsnAspAlaThrTyr 650
209 AAAGCCATCATGTTGGGCCACCTGATGATTCACAGCGCGCTTAAGA 258
650 ..... 650
259 AATATCGCTACATGTCGCTTTCCGATCAGGGCAACAAATTCATTC 308
651 .....ValProIleThrThr 656
309 GCCCTTCGACAAACCATGCTCAC...ATTGCGATTTGACGAAAGCGGTA 355
657 ...ProGluThrThrIleGlnSerGlyLysProThrPheAspLysProAsp 672
356 GTCCCGTTACGAGATTCAGCTTTAC..... 382
673 ValProLeuAlaAsnAspAlaPheSerValLeuAspValTyrAsnLys 689
383 .....GCATCCATTCGAGCGATACGAAC.....ACCATCCGCGC 416
689 pHeGlyAsnAlaSerValAspAlaAsnThrGlyLeuValThrPheThrP 706
417 CGAGCGCTATGACGGGCCACAGG.....GCGCGGCTATCCGCTC 457
706 roAlaLysGlyValGlyGlnSerGluProIleThrGlyThrIleProIle 722
458 CCAAGGCGGAGGATATATACCTAGACATTAAGCGGTGCGCA 507
723 ...LysIleValTyrGlnAspGlySerValGlyThrThrAspLeuAlaVal 738
508 AATATCGCTCACCTGACCGCAACCGCAGCCGACGCAACGGCTTGC 557
739 ThrValSerLysAsnIleTyrGluAsn.....ProGlyGluAsnIlePr 753
558 CGACCGTTCCCAATGCC.....GGCGCATGCTCA 589
753 oAlaGlyTyrHisLysValThrPheThrAlaGlyGlnGlyThrSerIleG 770
590 CGCAAGGA.....GTAGCGGACGATTCAAACGCGCGCAC 624
770 IuSerGlyThrThrValPheAlaValLysAspGlyVal.SerLeuProG 786
625 CGATACAGCCCG.....AGCT 641
786 uAspLysLeuProValLeuLysAlaLysAspGlyTyrThrAspAlaLys 803
642 GGACAGATCGGCAATGCCCGCA.....GCCTTCAC 675
803 rProGluGluAlaThrGlnProIleLysAlaAspThrGluPheVal 819
676 GGCACATCATATCGTCAAAACATCATCGCGCGCAGAGAAATTC 725
820 SerSerAlaThrLysLeuAspAlaIleIleGluAsnProGlyLysAsn 836
726 CGGCGCAGGCGATGCCGTGACGGGTATACGGAAGGCTCAACATTGCTG 775
836 eProAlaGlyTyrHisLysValThrPheThrAlaGlyGlnGlyThrSer 853
776 TCATGACGCGCTTGCTGCTTCCACGCAAAACAAAGATGGCGCGCAT 825
853 IeGluSerGlyThrThrValPheAlaValLysAspGlyValSerLeuPro 869
826 AACGATTTGGCAGATATGCGCAACTCAAGAC...TATCCCGCAGCAGC 872
870 GluAspLysLeuProValLeuLysAlaLysAspGlyTyrThrAspAla 886
873 CATCCGGAATTTGGCAGTCCAAACCCCAATGCCGACAAAGCATAGAAC 922

```

```

886 sTrProGluGluAlaThrGln...ProIleLysAlaAspThrGluP 902
923 CCCTGACCAATATCTTATGCGACGCATCCCATCAAGGATGGA... 969
902 heValSerSerAlaThrLysLeuAspAspIleIleGluAsnProGlyAsp 918
970 .....GCTGTCGGGGAA 983
919 AsnIleProAlaGlyTyrHisLysValThrPheThrAlaGlyGlnGly 935
984 ATACGCTTGCGGCGCATCAGCGCACATCTCTCAAGCGGTGCAATGG 1033
935 rSerIleGlnSerGlyThrThrValPheAlaValLys.....A 948
1034 GCGGATCGCATGCGCAAGGAAATCCGCGTC.....ACGAC 1074
948 spGlyValSerLeuProGluAspLysLeuProValLeuLysAlaLysAsp 964
1075 AATTTGCGCATGCGCGCATACGCCAAATACCGCTCCCTTACCATTCGG 1124
965 GlyTyrThrAsp.....AlaLysTrpProGly..... 973
1125 AATATCCCTTCAACCTTGAGACGCTTACGGCAAGAAACATCACC 1174
974 .....G 974
1175 CCTCACCGTCCGCGCTCAACGCAAAATGTCAAACTGCGACGACCA 1224
974 IuAlaThrGlnPro.....IleLysAlaAspAspThr 984
1225 CGCCACCGGAAGACAGCGCTACCGTTGACGCTGAAGGTTCCGAATT 1274
985 GluPheValSerSerAlaThrLysLeuAspLysSer....AspAl 999
1275 TGAGAGACGCTG.....AATATGATACGACGCTGATTC 1312
999 AspLysTyrThrProGluGlnLysValThrThrGluLeuAsnLysG 1016
1313 AAGATTTATCGGGGGCGGTATACCTAAGCTAAGCTGTG...TTTGAT 1359
1016 IuProAspAlaSerGluGlyLysLysAsnLysLysAspLeuProLysAsp 1032
1360 CGCAACCGCATG.....GAGGTGATGAGAACTTATTA 1397
1033 AlaLysTyrThrLysGluLysValAspIleSerThrAlaGlyAsnLys 1049
1398 ATTGACACT.....CGTGACGAGGTG 1420
1049 sLysGlyThrValValThrTyrSerAspGlySerSerAspGluValG 1066
1421 AGAAATATGTCAGGAAACGAGAAAGAGCTCAGACTGACTTAA 1470
1066 IuValAspValThrValThrAspAsnArgSerAspAlaAspLysTyrGlu 1082
1471 GCCATGCGCAACGAAATGGGAAATTAACAGGTTAGTTTAA 1520
1083 ProThrValGluGlyGluLysValGlu..... 1091
1521 TTTTATAGTGTGATATCATTAAGAAAGCACATACAGAGGAGATA 1570
1092 ...IleGlyLysValAspLeuThrAspAsnValThrAsnLeuPro 1107
1571 GTCTAACCGTGT.....GATGTACGGGTG.....ATACAA 1602
1107 hrLeuProGlnGlyThrThrValThrAspValThrProGlyGlyThrIle 1123
1603 CAACCTCGGACCTGATTAACATGCGTTATATCAAGCAGAGGGAAT 1652
1124 AspThrAsnThrPro.....GlyAsnTyrGluGlyValIleGluVal 1137
1653 TAAAAAGCCTGATGGAAGTGGAGGTGAAACGAAAGAGGTGGGAAG 1702

```



```

302 1eSerSerThrlleSerValGlnProSerSerSerSerSerProthr 318
989 GCTTGGGCGGCATCAGCGACATCTGTCAAGCGGTCCAGATGGCGCG 1038
319 ThrSerThrlleSerValGlnProSerSerSerSerSerProthr 335
1039 ATCCG.....CATGGCCAAAGGAATCCGGCGGTAGGGA 1073
335 rSerAlaThrSerValGlnProSerSerSerSerProthr 352
1074 CAATTTGGCGATGCGCATGCGCAATACCCGTCCCTCCATTCATCC 1123
352 eThrThrleSerValGlnProSerSerSerSerSerProthr 368
1124 GAAATTCGCTTCAACTTGGAGCAGCTTACGCCAAGAAACATCACC 1172
369 ThrThrSerValGlnProSerSerSerSerSerProthr 385
1173 .....TCCTCACGCGTCCGCGGTCAACGCA 1201
385 aThrSerValGlnProSerSerSerSerSerProthr 402
1202 AAATGTCAACTGCGACACACCGACCGAGACGCGGTACCGTT 1251
402 hThrSerValGlnProSerSerSerSerSerProthr 416
1252 GACGTTAAAGGCTTCCGATTTTGAGAACGACGTGAATGATACGA 1301
417 ThrThrSerValGlnProSerSerSerSerSerProthr 433
1302 GCTCGATTTCAAGATTAATTCGGGGCGGTATACCT 1338
433 aThrSerValGlnProSerSerSerSerSerProthr 445
seq_name: sp_invertebrate:q9yU43
seq_documentation_block:
ID 09yU43 PRELIMINARY: PRT: 954 AA.
AC 09yU43:
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DE 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE CG11274 PROTEIN.
GN CG11274.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132.
RA Adams M.D., Gelinkker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abdril J.F., Adiyanti A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA dePablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

```

```

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Switskay R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003539; AAF49848.1;
DR Flybase: FBgn0036340; CG11274.
DR InterPro: IPR002483; PWT.
DR InterPro: IPR000230; Ribosomal_S12.
DR Pfam: PF01480; PWT; 1.
DR SMART: SM00311; PWT; 1.
DR PROSITE: PS00055; RIBOSOMAL_S12; UNKNOWN_1.
DR SEQUENCE 954 AA: 107592 MW: 57221697910A107 CRC64;

```

```

alignment_scores:
Quality: 145.00 Length: 646
Ratio: 0.495 Gaps: 33
Percent Similarity: 45.356 Percent Identity: 22.136

```

alignment_block:

US-09-303-518D-463 x Q9YU43

Align seg 1/1 to: Q9YU43 from: 1 to: 954

```

66 ACACGCTCAGATTTGGCAACGATCCCTTATCCGCGAGTCTGCACC 115
398 SerArgSerArgArgLeuSerArgSerProMetArgTyrSerArgSerPr 414
116 GTCACCATTTGAAACCGCAGCGGAAATACCACTATTCGAGACGGGG 165
414 OArgArgPheAsnAsnArgArgArgSerProMetMetAsnPheArgLys 431
166 GAGCTTGGCCNAGCGCAACGCGCATATCGATTGGAAACATACAAACCA 215
431 LysLys..... 432
216 TCAGTTGGCCACCTGATGATTAACAGCGCGCGTGAAGAAATATCG 265
433 .....ArgGlyGlyLysArg. GlyLysLys..... 440
266 GCTACATTTGCGCGTTTCGATCAGCGGCAAAATTCATCCGCCCTTC 315
441 .....HisArgLysMet 444
316 GACAACTAGCTGCATTCGATTCGACGAGCGGCTAGTCCCGTTGA 365
445 TTrGlnHisArgGlyGlySerProAsnPheArgLysLys..... 458
366 CGGATTCACGCTTTCCGATTCATTTGGAGAGGATACGACACCATCCG 415
459 .....ArgLysHisTrpGlnLa...ArgHisSerProG 465
416 CCGAGCGCTATGACGCGCACAGGCGGCGGTATCCGCTCCCAAGGC 465
465 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 485
466 GCGAGGATATATACAGCTACGACATAAAGCGCTTCCCAAAATATCCG 515

```

485 rgGlnSerProGlnGlnAsnArgTyrAsnArgSpHisArgGlnSerPro 501
516 CCTCAA.....CCTGACCGCACACC 535
502 MetGlnGlnAsnGlnProPheArgTyrGlnPheSerProHisGlnGlyPro 518
536 GCACGACCGGACACAGCGCTTGCCGACCGCTTTCGACCAATGCGCGCGGTATG 585
518 ocGlyArgArgPheSerSer.....ProGlnGlnArgArg.... 529
586 CTGACCGCAAGAGACTGAGCGACGGATTCAACGGCGCACCGCATATACACCC 635
530 ..AsnSerArgSpArgArg..ProAsnSerArgGluArgArgSerSer 545
636 C.....GAGCTGGACAGATGGCGCATGCCCGCCGAAACCT 670
545 ocGlyGlyGlyGlyGlyHisMetAsnArgTyrPaspAsn..ProProProAla 561
671 TCAACGCGACTGCAGATATGTCATAAAACATCATCGCGCGCGAGAGAA 720
562 ArgAsn.....ArgArgSerSerSerGlySerSerAla..GlyGlyArg 576
721 ATTGTGCGGCGACCGCATGCCGTGCAGGGTATATAGCGAAGCGCTCAACAT 770
576 GlnGlnArgGlnArgSerAlaSerProProGlnLysArgArgSerHis 592
771 TGC.....TGTCAATGACAGCGCTTGCGTC 793
593 SerArgSerArgSerLeuSerArgAsnArgSerArgSerArgSerSer 609
794 TGTCTTCCACCGAAACAGATGCGCGCATCAACGATTGGCAGATATG 843
609 rGlnHisSerArgLysArgGlnSerPro.....IleGlyArgSerS 623
844 GCGCA.....ACTCAAGACTATGCGCGCAGC 869
623 erValGlyTyrAlaGlyProAlaValAsnThrIleAspLeuLysArgGlu 639
870 AGCCATCCGGGATTGGCGACTCCAAA..... 896
640 GlnGlnGlnArgLysAsnLysProGlnThrIleIleValIleuAspGln 656
897 .CCCCATGCCGACACAGCAT.....AGACCGCGACGAT 933
656 yProAlaValArgSerSerTyrAlaSerLeuSerArgThrProSerPro 673
934 ATCTTATATGGACAGCATCCCGCATCAAGGAGTGTGCGCGGAAA 983
673 heLeu.....LysProHisGlnArg..LeuAlaAla..... 682
984 ATACGGCTTGGGCGCATCAGCGACATCTGTGCAAGCGGTGCGAGATCG 1033
683AlaLysAlaAlaAlaAla 689
1034 GCGCGATTCGATTGCCGAAA.....GGAAATCCGCGCTC 1068
689 IAsnAlaAlaValProLysLysSerArgSpGlnSerGlnSerSer 705
1069 AGGACAATTTGGCGATGGCGATACGCCAATACCGTCCCTTACCA 1118
706 SerAspSerSerAlaGlnSerAspSerGlu..... 716
1119 TTCGCGAAATATCCGTTCAACCTTGAGCAGCGGTACGCGAAGAAACA 1168
717GlnGlnProArgArgLysLysGlnP 725
1169 TCACCTTCCTCAACCGTGCGCGCTCAACGCGAATAATGTCAACTGGCA 1218
725 roThrAsn.....MetGlnThrSerLysGlyLysGlnLysValArg 739
1219 GACCAACGCCACCGCAAGACAGCGATACCGTTTGACGCTAAAGGCTTCC 1268
740 ArgSerSerGlnSerSerThrSer..... 747

